Journal of Manufacturing Systems 63 (2022) 491–503

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

A graph-based reinforcement learning-enabled approach for adaptive human-robot collaborative assembly operations

Rong Zhang [a], Jianhao Lv [a], Jie Li [a], Jinsong Bao [a,*], Pai Zheng [b], Tao Peng [c]

[a] College of Mechanical Engineering, Donghua University, Shanghai 201620, China
[b] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China
[c] Institute of Industrial Engineering, School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Keywords:
Human-robot coexisting
Part-behavior assembly and/or graph
Behavior prediction
Self-attention
Adaptive decision making
Reinforcement learning

ABSTRACT

In today's prevailing manufacturing paradigm of mass personalization, neither human operators nor robots alone can perform all assembly tasks efficiently. To overcome it, human-robot collaborative assembly shows its great potentials to ensure the flexibility of human operations with high reliability of robot assistance. However, it is often challenging to achieve harmonious coexistence between humans and robots to complete the tasks safely and efficiently. In this regard, this research provides a detailed description of the human-robot coexisting environment and further introduces key issues in collaborative assembly. A part-behavior assembly and/or graph based on process requirements is proposed to represent the assembly task of complex products. Moreover, the human behavior prediction network based on self-attention can achieve higher accuracy. Combined with the robustness of Soft Actor-Critic (SAC), the collaborative system improves the self-decision ability of the robot in the dynamic scene. Finally, the effectiveness of the method is verified through experimental analysis. The results indicate that the accuracy of the proposed behavior recognition based on self-attention method is 91%. At the same time, it is proved that the reinforcement learning method is theoretically feasible to provide adaptive decision-making for robots in human-machine collaboration. In addition, the convergence speed of the reward function proves the feasibility of SAC for adaptive decision-making in a human-robot collaborative environment.

1. Introduction

Human-robot collaboration is always a critical issue for intelligent manufacturing, especially for the assembly process of high-precision products [1]. The requirements of different assembly processes are different. For example, simple but laborious tasks are usually done by robots, while some complex tasks are assisted by robots [2,3]. Human-robot collaborative tasks in the coexisting environment require much interaction. For example, it is straightforward for humans to arrange tasks for robots through a controller. Correspondingly, it is complicated for the operator and the robot to interact and collaborate smoothly in the coexisting environment. Especially, the reaction strategy for the accidental contact should be distinguished from the activate collaboration since they tend to cause opposite consequences, which motivates the research on human intention recognition [4–6]. In general, a typical human-robot collaboration assembly consists of three procedures, the recognition and processing of assembly tasks, the understanding of human behavior and the calculation of space occupancy, and the updating of adaptive collaborative strategies for robots [7,8].

This paper focuses on adaptive work for human-robot collaborative assembly tasks. First, the robots are trained to identify and understand the intentions of human operators and then perform assigned collaborative subtasks. In this way, the efficiency of the collaboration and accuracy of the task execution can be improved [9]. A self-attentive mechanism approach is employed with a double-layer long short-term memory (LSTM) to extract behavioral categories from the human skeleton sequence feature. This approach enhances the recognition of human behavioral intent and improves the accuracy of human-robot collaboration. Accordingly, for the non-collaborative type of tasks, robots should actively stay away from the operator to avoid collisions. Therefore, understanding the relationship between human behavior and the assembly task is crucial for the robot to choose the corresponding execution strategy.

For such a problem, a representation method is proposed that integrates operation behavior into the assembly properties. Assembly and/or graph based on the product structure can increase the flexibility of the

* Corresponding author.
E-mail address: bao@dhu.edu.cn (J. Bao).

https://doi.org/10.1016/j.jmsy.2022.05.006
Received 3 March 2022; Received in revised form 23 April 2022; Accepted 7 May 2022
Available online 16 May 2022

assembly system and simplify the search process of the feasible plan in assembly processes [10]. Integrating the assembly behavior node to generate the part-behavior assembly and/or graph can link the assembly process requirements of the product with the corresponding operation behavior. Assembly information helps to refine the division of collaborative tasks. At the same time, behavior-based detection can verify whether the current tasks are accurate or not. Some scholars use the form of and/or graph to solve the task sequence planning problem in assembly [11]. In comparison, we adopt the SAC algorithm with learning ability to transform sequence planning and optimization into the process of learning and exploration.

Nevertheless, for different environments and task states, how a robot adjusts its strategy in time and completes collaborative assembly accurately and efficiently is the most critical aspect in human-robot collaboration [12]. Reinforcement learning methods are often used to represent the self-learning process of intelligent robots [13]. By actively interacting with the environment, the agent performs actions to update the environment state, receives rewards or penalties for changes in the state, and continuously updates its actions to achieve learning evolution [14]. Compared with a genetic algorithm [3], the SAC algorithm introduces entropy regularization, which can improve the randomness of strategy selection in the training process [15]. It can improve the robustness of the robot when applied in the human-robot cooperative system.

The contributions of this paper are summarized below.

1) An assembly task expression method based on part-behavior assembly and/or graph is proposed to improve the flexibility of collaborative tasks.
2) A behavior intention recognition network based on self-attention is proved to be beneficial to improve the accuracy of behavior prediction.
3) The reinforcement learning algorithm based on SAC is applied to the adaptive control of the robot in human-robot collaboration, which can enhance the robustness of the collaborative system.

The remaining content of this article is organized as follows. In Section 2, related work is presented while in Section 3, a human-robot coexisting environment is introduced, defining the components, the space, the state, and the control of the robot in the assembly environment. Section 4 introduces the construction of assembly and/or graph that fuse human behavior, the query, and the update method. Section 5 presents a self-attention network with the double-layer LSTM network to improve the accuracy of human behavior recognition. In Section 6, the enhancement and adaptive methods of human-robot collaborative assembly are described. The adaptive optimization method based on SAC will not only improve the learning efficiency of the robot, but also improve the robustness of the collaborative assembly environment. In Section 7, it is verified that the detection accuracy of the human intention recognition network is improved. Meanwhile, the reinforcement and adaptive network of collaboration are developed, which proves the effectiveness of the proposed scheme. Finally, the results are concluded in Section 8.

2. Related work

Research on human-robot collaboration strategies is numerous and has been ongoing for a long time [16,17]. In traditional manufacturing industries, assembly activities are time-consuming and energy-intensive, especially in an aging society [18]. In this regard, Raessa et al. proposed a human-in-the-loop robotic manipulation planner for collaborative assembly. The system distributes the subtasks of an assembly to robots and humans by exploiting their advantages [19]. In addition to structural design and precision control, the key technologies of human-robot collaboration mainly include semantic expression of unstructured information, human behavior recognition in dynamic interaction, and adaptive decision-making of robots in complex scenes.

2.1. Behavior intention recognition

In the process of human-robot collaboration, accurate identification of human behavior intention helps improve the accuracy of the assistant action of the robot. Ravichandar et al. presented a new method to infer human intentions denoted by the goal locations of reaching motions using a neural network-based approximate expectation-maximization algorithm with online model learning [20]. Lee et al. proposed that, captures different human motion patterns and the unmodeled dynamics, thus obtaining more accurate prediction [21]. Andrianakos et al. present an approach to automatically monitor the execution of human-based assembly operations using vision sensors and machine learning techniques [22]. Wang et al. introduced a Cyber-Physical-System (CPS) with a layered architecture, allowing robots to assist human operators with the expected motion estimation [23,24]. In the other case, the context-aware human-robot collaborative safety system can recognize human posture while doing path planning to avoid obstacles and still reach the target location in time [25]. Ding et al. proposed a data-driven programming approach in human-robot collaborative manufacturing systems to describe the human-robot interaction in each process [26]. In addition, Lin et al. applied the hidden semi-Markov model (HSMM) model to human-robot collaboration to predict human assembly efficiency [27]. In order to achieve natural interaction and enable robots to accurately and efficiently acquire information about the external environment, Zheng et al. studied hand recognition and human body re-identification in human-robot interaction, so that human attention can be accurately read and action intentions inferred [28–30].

It can be seen that in the research of human intention recognition, most scholars seek for evolution rules from the sequence information of human behavior, and speculate human intention based on the prediction method, which has a partial feasibility, but ignore that human behavior is not completely independent and often purpose-oriented, especially in industrial scenes. At the same time, due to the different flexibility of each joint of the human body, the frequently changing parts should be focused on, which is often the key factor to distinguish the intentions of different human actions.

2.2. Semantic representation

In terms of informationization description of assembly tasks, Faber et al. expressed all possible sequences by constructing a complete assembly of the product, and solved the optimal assembly sequence to meet the requirements by evaluating the number of human-robot interactions and whether each step is in line with ergonomics. The solution efficiency was significantly reduced for the product with a complex structure and more parts [31]. Furthermore, introducing self-optimizing and self-learning control is a crucial factor for cognitive systems. Bannat et al. integrated human workers into the workflow of a production system. This research described the worker guidance systems for manual assembly with environmentally and situationally dependent triggered paths on state-based graphs [32]. Similarly, Wang et al. proposed that a practical human-robot collaborative assembly system should be able to predict the intention of human workers, and they model the product assembly task as a series of human motions [33]. Sadrfaridpour et al. studied the adaptive speed control of robots and reduced the workload of human beings [34]. In a hybrid assembly cell with human-robot coexistence, Tsarouchi et al. proposed an intelligent decision-making method for the sequential allocation of tasks to the operator and the robot, and validated the method with the assembly of a hydraulic pump as an example [35].

Informative expression is the key technology to realize the transformation from traditional manufacturing to intelligent manufacturing. The unstructured information is transformed into computable

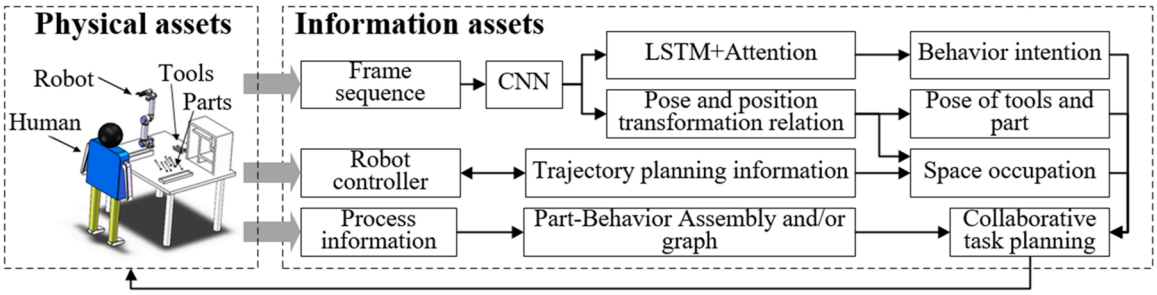


Fig. 1. Human-robot coexisting environment for assembly.

expression form, and the information of behavior, parts, tools, and task schedule in the operation process are fused to provide support for the intelligent decision-making process.

2.3. Self-decision making

Empowering robots with the ability to learn is an advanced presentation to achieve intelligence. By migrating manual operational tasks to a robot, the robot can continuously interact with its environment through reinforcement learning algorithms to maximize reward or achieve a specific goal [36,37]. A deep reinforcement learning model based on a policy search algorithm enabled the end-to-end manipulator control, avoiding relearning caused by changes in the spatial position of objects [38]. Loftus et al. combined reinforcement learning with the decision analysis for surgical applications in which the sequential recommendations evolved with changes in clinical conditions over time [39].

Further, to improve the adaptive capabilities of offline reinforcement learning, a bio-inspired hierarchical cognitive system is proposed for online learning and error correction of object concepts through human-robot collaboration [40,41]. It can help remember correct knowledge, quickly forget the false representation, and respond to human error effectively. Oliff et al. aimed to overcome the human factor of uncertainty involved in the human-robot collaboration. It improved the adaptability and autonomous decision-making ability of robots [42]. In summary, extracting knowledge of industrial manufacturing processes and enabling the transfer and processing of knowledge with human experience is a trend in human-machine collaboration.

Compared with the method of deep learning, the adaptive decision-making method based on reinforcement learning has more obvious advantages. Endow the robot with the evolutionary ability of self-learning, find more appropriate decisions in the continuous iterative optimization process, and effectively improve the robot's intelligence level to deal with different task situations.

3. Human-robot coexisting environment for assembly

The ultimate goal of human-robot collaboration is to communicate among the robot, assembly environment, and the operator, and achieve a harmonious state through the information interaction among them. It is easy for humans to understand the environment and the robot's compliant behavior because of the high degree of perception and cognition of human subjective consciousness. The cognitive ability of agents is often based on perception. For the robot, how to identify, understand human behavior intention, and obtain the state of the assembly environment requires additional information, such as human body position, tool position, tool posture, and so on. This information can be obtained by sensing equipment to enhance the perception ability of the robot.

To be able to create assembly plans and design the implementation process for these plans, in this Section, firstly, physical assets and information assets are introduced respectively to intuitively explain the information interaction process in human-robot collaboration (Section 3.1). The spatial layout of each element in the physical asset will be briefly described in Section 3.2. Secondly, in the task execution stage, the implementation of the assembly task is represented by the process of continuous interactive update of the operator and task state (Section 3.3). Finally, the flow of the robot collaborative control scheme calculated based on state information is given in Section 3.4.

3.1. Cyber-physical assets

Human-robot coexisting environment aims to realize the information and operation interaction between the entity objects, and jointly complete the target task under mutual assistance. As shown in Fig. 1, the elements of human-robot coexisting contain two parts: the physical and the information assets.

3.1.1. Physical assets

The physical part of the human-robot coexisting assembly environment consists of two main parts: the operating module (operating workers, collaborative robots) and the operating objects (assembly parts, operating tools).

In the operation module, the workers have high flexibility and strong decision-making ability, and can solve the flexible assembly works more effectively. However, humans have the characteristics of easy fatigue, and long working hours will lead to a lower willingness to work, poor stability in performing tasks, and a certain chance of making mistakes. On the contrary, a robot can guarantee repetitive movements with high precision, but it has relatively low adaptability. Therefore, the operation module formed by the complementary advantages of humans and robots can highly adapt to a variety of assembly tasks. The object module includes all parts to be assembled and various operation tools used in the execution of tasks. Each part of the assembly product is classified by main parts and connectors, and placed in the fixed area of the console.

3.1.2. Information assets

In a human-robot collaboration environment, the executor is required to obtain information and understand the environmental state for decision-making. The assistant robot usually needs additional auxiliary equipment to obtain external information. The most commonly used methods are visual, tactile and voice-based interaction. The architecture of human-robot-environment interaction based on visual sensing is shown in Fig. 1, divided into three parts. Firstly, from the image information, the human operation intention and the position and posture feature of other parts in the environment can be extracted, and the occupation in the workspace can be calculated. Secondly, based on the operational requirements and space occupation of the task, and the part behavior assembly and/or graph constructed based on the production process information, the robot generates the subtasks that need to be executed in the current state, and plans the safe and efficient cooperative assembly trajectory to realize the cooperative assembly task.

In the information interaction model of human-robot coexisting, on

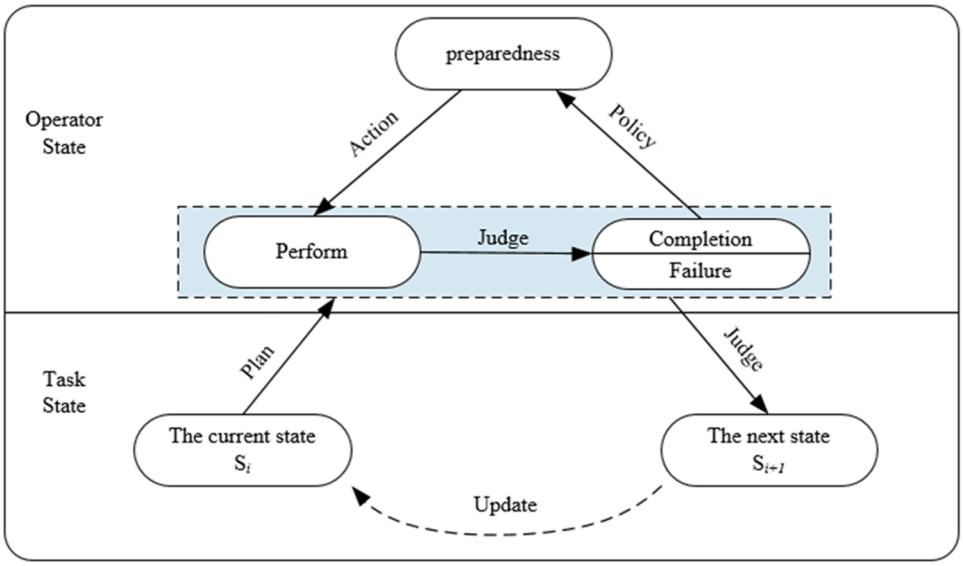


Fig. 2. Operator state transition and task state transition.

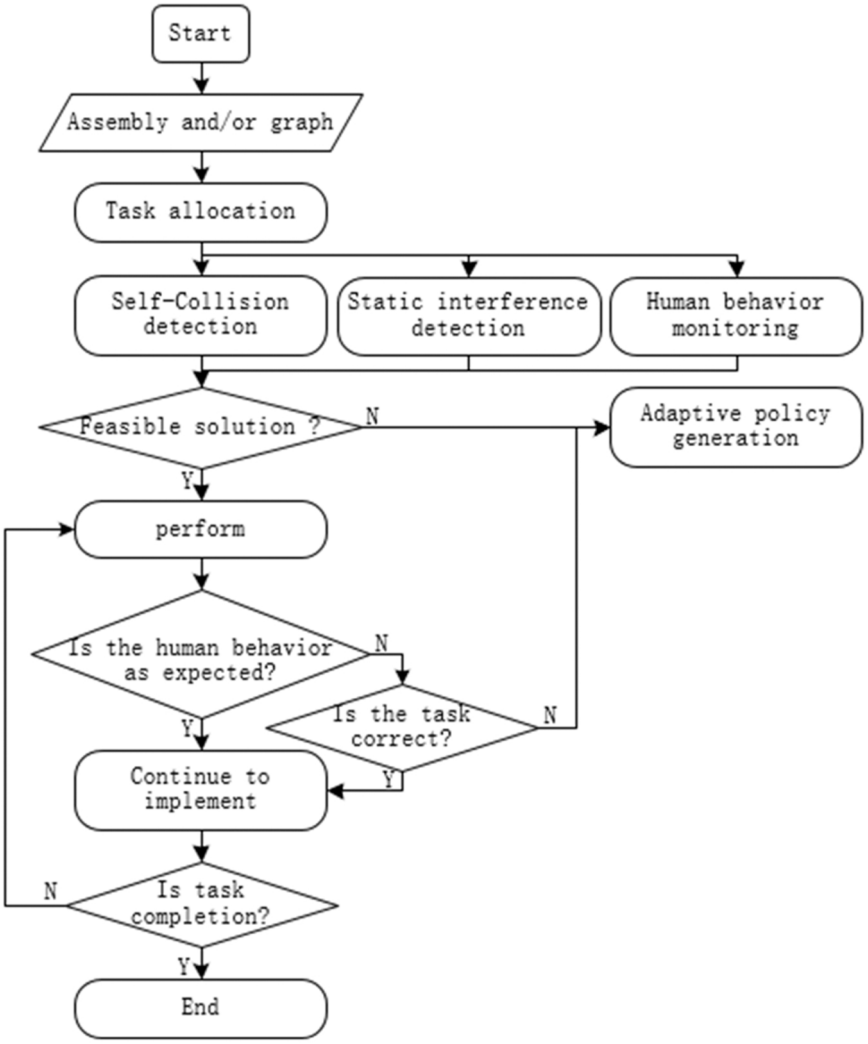


Fig. 3. The scheme flow of robot control.

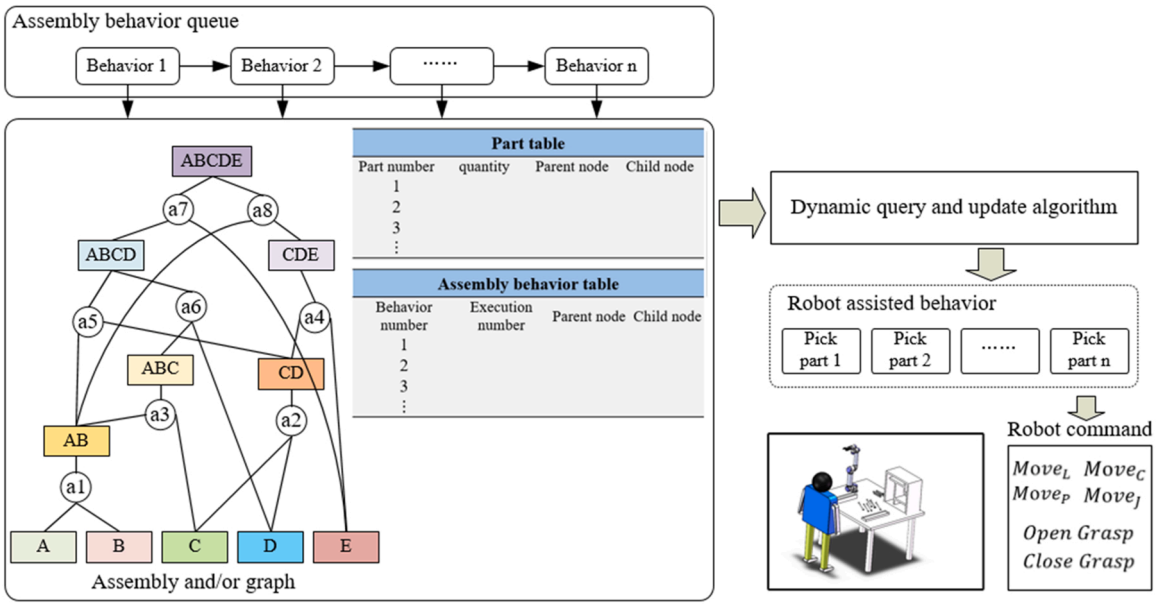


Fig. 4. Dynamic human-robot collaboration based on part-behavior assembly and/or graph.

the one hand, the robot inferred the human behavioral intention by processing the sequence feature of human posture and carried out the collaborative task under the human will. On the other hand, the robot recognized the state of the target object in the environment, judged the current task progress, the location of the operation part and other information, and carried out timely pickup and delivery to improve the task efficiency.

3.2. Spatial

The distinctive feature of the human-robot coexisting environment is the high degree of overlap between the human and robot workspaces. Considering human safety, a comprehensive analysis of the properties of various space components is particularly important for the harmonious coexistence of humans and robots. The collaborative assembly workspace under human-robot coexisting is mainly divided into three parts: human workspace, collaborative workspace and robot workspace.

The human workspace can be divided into two parts: the comfortable operating space and the reachable space. For the comfort space, the human operation is efficient. For the reachable operation space, humans are required to put more energy and more likely to cause fatigue. Therefore, in the environment layout, the places and tools are classified based on the abilities of humans and robots to reduce extra workloads as much as possible. The safe collaborative space is the intersection of human and robot workspace, usually located at the operator's desk. In the collaborative space, difficult tasks are usually performed that require the participation of both humans and robots. The robot operation space is usually determined by the rotation angle limit of each joint of the robot and the length of each robotic arm, as a sphere-like operation space. Considering the flexibility of robot operation, it can be further divided into the reachable space and the flexible workspace. The robots prefer to perform complex collaborative tasks in a flexible workspace which facilitates planning and enhances the flexibility of collaboration.

3.3. State

Similarly, according to various properties, the state can be divided into the state of the executor and the state of the execution object. As shown in Fig. 2, the executor's state usually includes preparation state, execution state and result state (completion and failure), and changes with each other and circulates continuously until the task is completed. The task state of the execution object can be understood as the execution progress state of the current subtask, that is, from the initialization state to the final assembly completion state after the execution.

3.4. Control

A reasonable and practical robot control scheme can greatly improve the efficiency of human-robot collaboration. As shown in Fig. 3, first of all, input the part-behavior and/or graph of the product to be assembled. The part properties information and assembly process requirements extracted from the part-behavior and/or graph, assign assembly tasks to operators and robots in a reasonable and orderly manner. Then, based on the task requirements, compute the feasible solution without collision, and the robot starts to perform the corresponding assembly operation. In the process of robot movement, in order to judge whether the execution of collaborative tasks is in line with the intention of humans, it is necessary to recognize effective operation intention through human behavior, and judge whether human intention changes, to verify the rationality of robot operation behavior. If the human behavior intention changes, we need to further judge whether the current task is correct, and generate effective solutions through an adaptive strategy algorithm. The scheme design of adaptive strategy will be introduced in detail in Section 6.

4. Task representation based on the assembly process

4.1. Part-behavior assembly and/or graph

In the manual assembly mode, there are many problems in how workers gradually assemble according to the assembly drawings. For example, the assembly route is not unique, the interference between the front and back assembly process and so on. The process of disassembly and reassembly will significantly reduce the assembly efficiency. As a representation method of dynamic assembly sequence, assembly and/or graph can update the current assembly state, and workers can obtain the following feasible assembly scheme through the query mechanism. Based on the assembly and/or graph, the assembly behavior node is integrated to form the part-behavior assembly and/or graph. As shown in Fig. 4, a part-behavior assembly and/or graph can be constructed based on the queue of assembly behaviors. The dynamic query and the update method can update the current assembly status, and provide

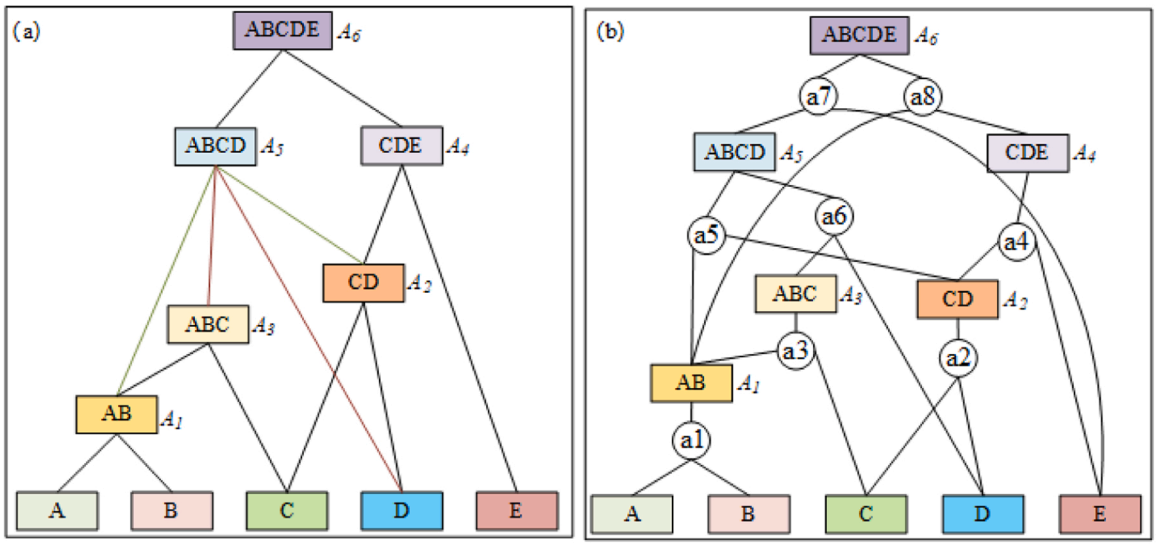


Fig. 5. (a)assembly and/or graph. (b)part-behavior assembly and/or graph.

Table 1
Properties of the part.

Part number	Quantity	Parent node of assembly behavior	Child node of assembly behavior
A_1	q_1	a3, a5	a1
A_2	q_2	a4, a5	a2
A_3	q_3	a6	a3
A_4	……	a8	a4
A_5		a7	a5, a6
A_6		None	a7, a8
A	q_i	a1	None
B	……	a1	None
C		a2, a3	None
D		a2, a6	None
E	q_n	a4, a7	None

Table 2
Properties of the assembly behavior.

Assembly behavior number	Number of executions	Assembly component	Assembly body
a1	e_1	A, B	A_1
a2	e_2	C, D	A_2
a3	…	A_1, C	A_3
a4		A_2, E	A_4
a5	e_i	A_1, A_2	A_5
a6	…	A_3, D	A_5
a7		A_5, E	A_6
a8	e_n	A_1, A_4	A_6

auxiliary operation for workers according to the feasible assembly sequence. The auxiliary behavior of the robot can be transformed into the corresponding robot control instructions to drive the robot by TCP/IP communication.

Specifically, assembly information is divided into three categories: the part information, the part relationship, and the assembly action. The part information contains its name, designation, and quantity. The relationship information of parts is expressed as the matching mode between parts, including screw fastening, peg-in-hole assembly, snap-fit, gluing, etc. Assembly action information represents various operation behaviors of the worker to part or subassemblies, which can be divided into selecting parts, adjusting parts, matching parts and fastening parts.

The part-behavior assembly and/or graph is transformed from the assembly and/or graph, and the behavior node is added, as shown in Fig. 5. Starting from the leaf nodes (parts) of the tree, the intermediate nodes (subassemblies) are generated continuously in a certain order, and finally assembled into the root nodes (final assembly). In the part-behavior assembly and/or graph, the subassemblies are generated dynamically according to the assembly behavior of workers. For example, in the assembly process of part A and part B, subassemblies {A, B} can be obtained by identifying workers' assembly behavior *a1*. In the assembly process of subassemblies {A, B}, subassemblies {A, B, C} and {A, B, C, D} can be obtained respectively according to assembly behaviors *a3* and *a5*. According to the updated result of the assembly state, the corresponding assembly process can be queried in the and/or graph, which can help the robot to complete the decision.

4.2. Storage structure

Due to the complex information involved in the part-behavior assembly and/or graph, a specific storage structure is needed to describe the assembly information effectively. Compared with Excel, MySQL database can store more assembly data, which is more suitable for storing and managing and/or graph information. Firstly, the part table and behavior table are generated in the database, and the part table information is defined from four quantities: part number, quantity, parent node and child node of assembly behavior.

As shown in Table 1, quantity refers to the number of corresponding parts under the same assembly task. For example, six identical nuts are required for fixing the rear end cover of the reducer. The parent node is used to query the possible subsequent assembly behavior of the current component, and the child node is used to query the possible assembly behavior of the current subassembly. According to Table 1, the parent node and child node of assembly behavior are not unique, indicating that there may be multiple feasible assembly schemes.

As shown in Table 2, the assembly behavior information table contains four parts, assembly behavior number, number of executions, assembly component and assembly body. The number of executions reflects the number of assembly operations for the same parts. For example, fixing the reducer rear end cover with nuts needs to be repeated 6 times. The assembly component represents two objects that execute assembly based on recognizing the current assembly behavior.

4.3. Query and update

Based on the part-behavior assembly and/or graph, the assembly sequence can become dynamic to meet the adaptability of human-robot

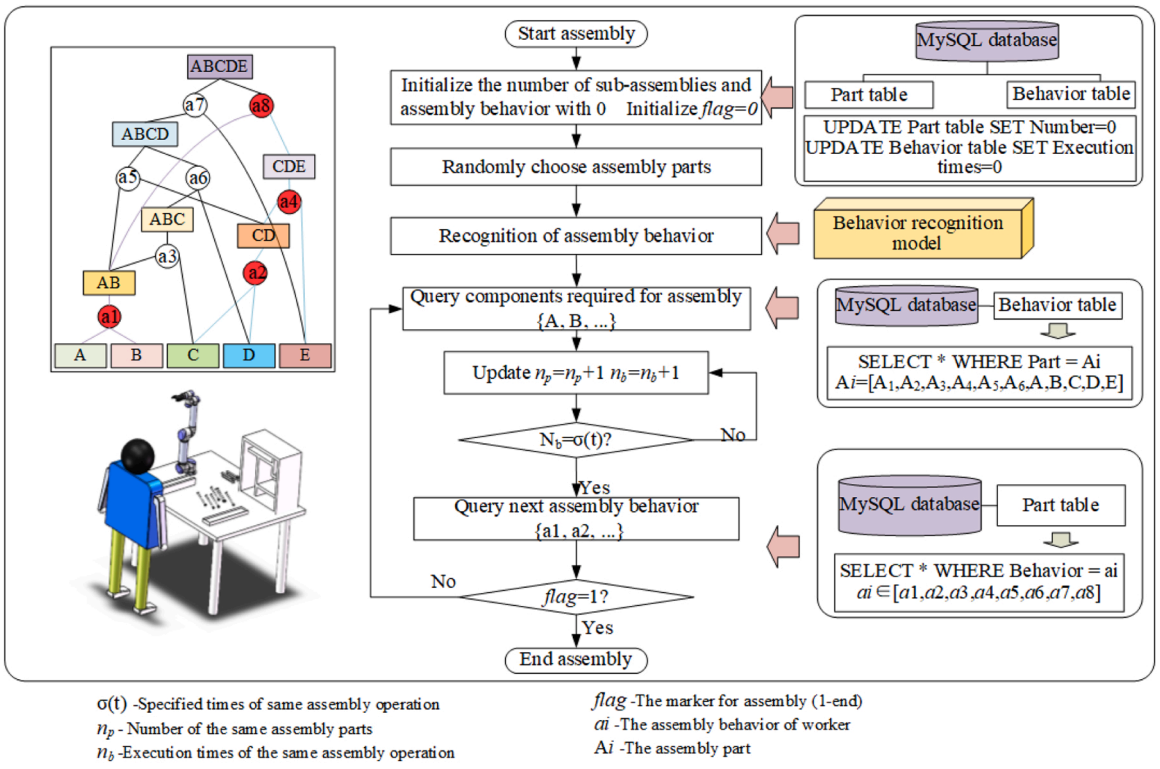


Fig. 6. Dynamic query and update method.

collaboration. As shown in Fig. 6, when assembly starts, the number of subassemblies and assembly behaviors are initialized to 0, using "UPDATE… SET…" command can set the corresponding columns in the part table and behavior table to zero. For the assembly with n components, workers can randomly select two components with assembly relationship to carry out the initial assembly. The assembly behavior of workers in the assembly process will be identified, and the parts needed for the behavior will be obtained in the behavior table through the query function of MySQL database. In addition, the MySQL database query statement used is "SELECT * WHERE Part = A_i". Some assembly operations of parts need to be repeated several times, and the number of parts in the part table and the assembly behavior times in the behavior table need to be updated. When the number of assembly operations executions meets the task requirements, the following feasible assembly behaviors will be queried in the parts table through the database. Moreover, the corresponding MySQL database query statement is "SELECT * WHERE Behavior = a_i". The dynamic query and update function of MySQL database will run to the end of the assembly process until the flag= 1.

5. Behavior intention recognition network based on self-attention

Due to the continuity of body behavior, LSTM is used to mine the action sequence information between adjacent frames and analyze human action behavior. The double-layer LSTM can analyze different features respectively to improve the accuracy of behavior recognition. In addition, the attention mechanism is integrated into the double-layer LSTM network to enhance the accuracy of human action intention recognition under the condition of the same behavior characteristics but different tasks. The network takes the continuous action frame as the input and the output is the task type performed by the operator.

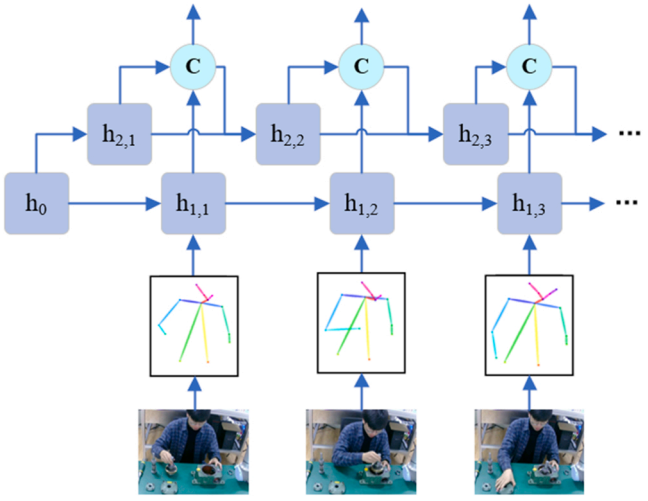


Fig. 7. Pose feature fusion of double-layer LSTM.

5.1. Double-layer LSTM

Each part of the human body is organically connected by joints, so the movement of the parts has a certain continuity. Although human behavior can be regarded as random, human posture changes are not discrete. Especially when workers are working, their action process will follow a certain procedure to a certain extent. There must be some connection between the decomposition actions of the operation process, which belongs to a "semi predictable sequence".

Although some typical human postures can be predicted by the skeleton feature in a single frame, for general human postures, the skeleton feature often shows a high degree of similarity, which leads to the distortion of semantic information. Therefore, it is necessary to extract the pose data from the adjacent frames by considering the

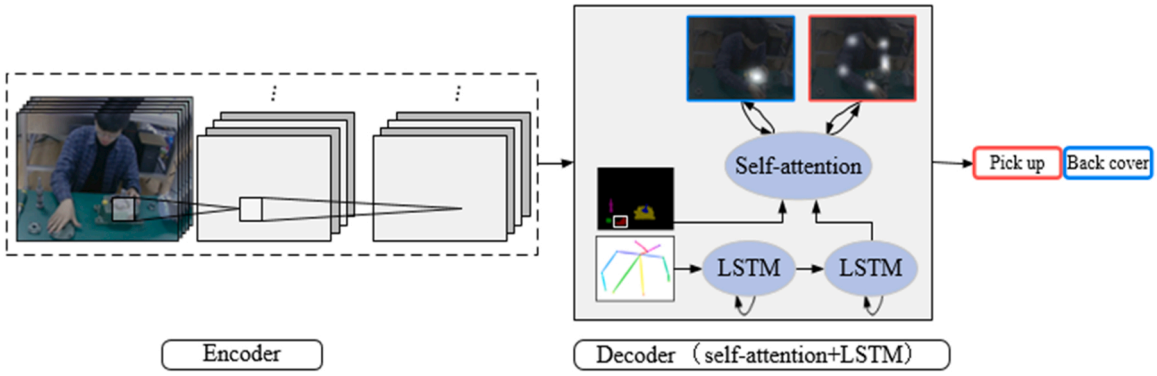


Fig. 8. Self-attention based sequential feature extraction framework.

sequence feature of the pose feature. Although the LSTM network has an absolute advantage in processing sequential features, it needs to add another layer to the traditional LSTM to process other information besides sequential features based on the high-dimensional complex sequential task security discrimination involved in this paper. Therefore, the double-layer LSTM is used to obtain information between different dimensions. The double-layer LSTM framework for sequential association mining of the human skeleton posture feature is shown in Fig. 7. The image at each moment is input to the network according to the time sequence, and the existing OpenPose model is used to extract skeleton information from images. The lower LSTM is used to analyze the high-dimensional skeleton feature online, and the upper LSTM is used to mine the internal relationship in the change of sequential features.

Firstly, the pose feature Pt of the current frame is input into the network according to the time sequence, and combined with the pose feature vectors of the last T moments. It is input into the LSTM of the bottom layer. The LSTM of this layer parses and expresses the pose feature of the human body. At the same time, the structure of this layer can also carry out a certain degree of short-term temporal correlation learning.

The upper LSTM is used to mine the long-term and short-term association relationship of the whole pose information sequence. The input of each time is the high-level feature of the previous time after feature fusion. It has low dimension and strong interpretability, and is suitable for context-related information modeling.

The state information corresponding to double-layers of LSTM is input into the fusion unit, and the final operation behavior characteristics are obtained after processing. Eq. 1 shows how to combine, where W_1 and W_2 are the combined weight parameters of the state output of each layer of the LSTM unit, which are generally obtained through synchronous learning in the training process.

$$C_t = W_1 h_{1,t} + W_2 h_{2,t} \tag{1}$$

For the input layer with complex features but relatively simple correlation within the sequence, the network will focus on the analysis of pose information at the bottom. For the input layer with more complex correlation than pose information, the network will focus on the analysis of complex dependence between sequences, and pay less attention to the interpretation of pose. Through this automatic selection mechanism, the model can choose the focus of network learning by itself in the training process to avoid excessive analysis of "useless" information and improve the flexibility of the network.

5.2. *Double-layer LSTM with self-attention*

However, behavior detection based on skeleton sequence features

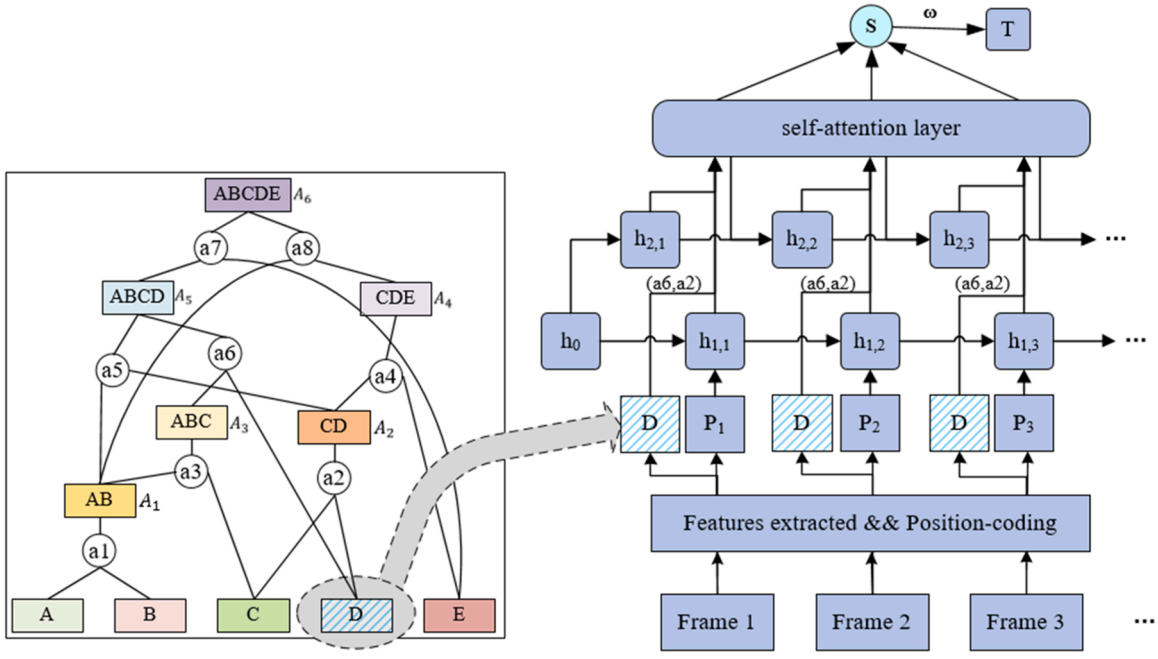


Fig. 9. Double-layer LSTM with self-attention.

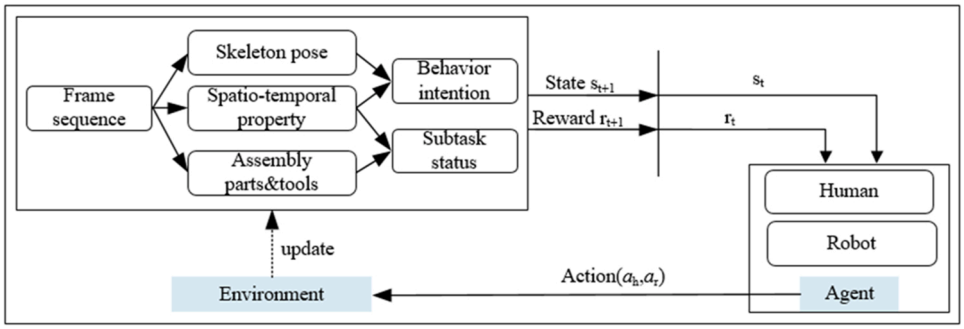


Fig. 10. Reinforcement learning based on human intention.

ignores the environmental factors on human actions and only focuses on the human posture sequence feature, which has some defects. Especially in the assembly scene, human perform operations frequently, and almost all behaviors are product-oriented. It is challenging to distinguish such many similar instantaneous postures or short-term states. Therefore, human posture features and other information besides sequence features are required to further enhance the accuracy of the pose prediction network.

By deeply analyzing the behavioral sequence information of human body with the information of human hands and parts in the images, the human skeleton information as well as detailed features of the hand, including gesture changes and the accompanying object state of the hand, are extracted to enhance the estimation results of human intention. As shown in Fig. 8, firstly, the features of human body are obtained based on Convolutional Neural Networks (CNN), and then the part information is extracted. At the same time, the human skeleton information obtained based on OpenPose is input into the double-layer LSTM network. Finally, the human behavior intention in this state is predicted by fusing with the part information in the self-attention layer.

As shown in Fig. 9, for each frame in the video stream, the historical skeleton feature sequence is directly obtained by querying the recorded frames in the memory. The skeleton feature of each frame is used as the input of a moment, and the input image information is added with the position-coding information through the entire connection operation of the feature extraction layer. Then, the skeleton feature and operation part feature in the image are extracted respectively. The parts feature and the deep skeleton feature obtained from the upper layer are fused and input to the self-attention layer to obtain the output features of each behavior. The output feature S is spliced and multiplied by an output matrix w to obtain the final sequential feature T.

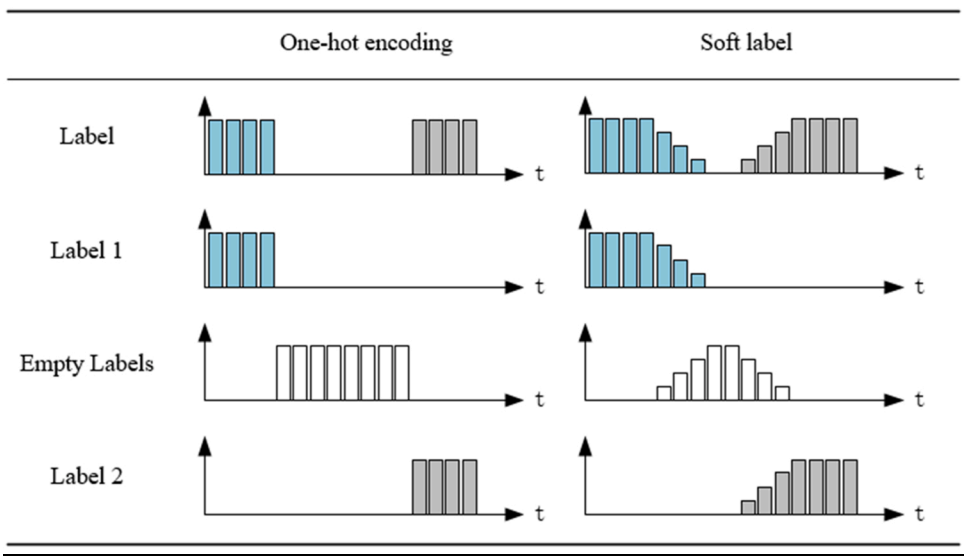
6. Reinforcement and adaptive method of human-robot collaboration

According to the recognized human behavior, the robot agent obtains the assembly parts to be operated simultaneously and finds the current assembly subtask based on the part-behavior assembly and/or graph. For the known subtasks, the robot needs to perform corresponding operations, such as grasping the tools needed by the operator, or assisting the operator to fix the base of the assembly. Robots often encounter situations where the expected action path is blocked during a task, so multiple optional alternate strategies are needed. As shown in Fig. 10, based on the control method of the reinforcement learning framework, the robot can continuously interact with the environment and the task with the guidance for new effective solutions by rewards.

In the reinforcement learning environment for human-robot collaboration, the state model should include the human state S_H, the robot state S_R and the completion progress state S_E of the assembly. The state of a robot is defined directly by itself, while the state of a human is obtained through intentions. Taking the operational state S_R of the robot as an example, it can be expressed as Eq. 2.

$$S_R = \begin{Bmatrix} (0, i+1) & \text{finish, select next target} \\ (1, i, T-t) & \text{assembly in progress} \\ (-1, i) & \text{waiting for collaborators} \end{Bmatrix} \quad (2)$$

Table 3
Comparison of the one-hot coding and the adaptive soft table.



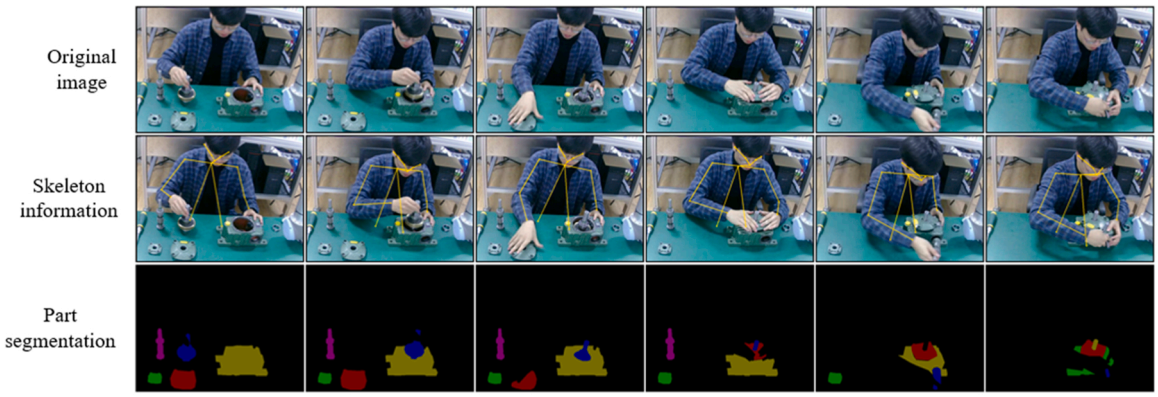


Fig. 11. Parts of the dataset.

In the equation, t represents the time spent in executing the current subtask; $T-t$ denotes the remaining time, as one of the factors that affect the agent's decision on collaborative tasks; i represents the part number.

The completion progress state is defined as $S_E = \frac{n}{N}$. n represents the number of tasks currently completed, and N represents the total number of tasks, which is equal to the total number of assembly parts. With the growth of S_E, it can effectively guide the agent to complete all the assembly tasks.

To guide the robot to learn the correct strategy faster, the final design of the reward function is shown in Eq. 3. Where M_{max} is the maximum number of steps allowed for one episode, if the number of steps is exceeded, this episode is considered a failure and ends immediately. J_{min} is the sum of the change ranges of each joint of the robot. D_p and D_b denote the linear distance of the robot from the target part and the obstacle, respectively. $\cos(o_H, o_r)$ represents the angle between the robot's movement direction and human, which provides comfortable avoidance space for humans. r_{end} is the reward in the last of an episode.

$$r = \frac{-\cos(o_H, o_r)}{M_{max}(D_p - D_b)} - \frac{1}{J_{min}} + r_{end} \tag{3}$$

In addition, to maximize the cumulative rewards, a reinforcement learning algorithm called SAC is developed based on the principle of maximum entropy. It maximizes the entropy of the policy while optimizing it for higher cumulative rewards. Integrating the SAC-based reinforcement learning into robot control systems for the human-robot collaboration, the agent can explore the state space and avoid the policy falling into local-optimum early. Furthermore, the multiple feasible solutions can be investigated to accomplish the specified task and improve the resistance to interference.

7. Experiment and analysis

7.1. A case study of behavior recognitions

7.1.1. Experimental setup

In general, the label of the classification model in training is in the form of one-hot coding. A multi-classification task is labeled as a one-dimensional vector of length N. The cross-entropy loss is calculated through the prediction probability and the real probability. Then the variable parameters of the whole model are updated through the gradient back propagation algorithm. However, the classification of sequential behavior is continuous and transitional, and the discrete label method fails to distinguish the behavior states at different times in the same sequence. For example, there may be meaningless behavior between various assembly behaviors or transitional behavior similar to the previous behaviors. In this paper, an adaptive soft label method is used to label video samples frame by frame. The comparison of the one-hot coding and the adaptive soft table is shown in Table 3.

Table 4
Comparison of the different recognition methods.

	Accuracy	Precision	Recall	F1-Measure
Ours	0.91	0.93	0.89	0.90
CNN+LSTM	0.77	0.81	0.74	0.77
ST-GCN	0.70	0.72	0.67	0.68

In this experiment, the RealSense camera is fixed on the front and top of the human workers, so that clear pictures of the workers' body behavior and the assembly state of the parts can be taken. The image resolution is 640 × 800, and the frame rate is 30 frames/second. According to the sequence relationship of part behavior and/or graph, 200 groups of video data are obtained. Each video is 16 – 24 s long and contains multiple different actions for assembling a particular object.

Then, the assembly behaviors of workers in the video data are labeled. Specifically, the obvious assembly behaviors are labeled according to the one-hot coding, while the uncertain behaviors with excessive characteristics are labeled according to the way of soft label. A total of 200 sets of frame-by-frame labeled video sequences of assembly behaviors are obtained, and parts of the dataset are shown in Fig. 11. Finally, the dataset is augmented to 400 groups of labeled videos using the horizontal flip method, of which 320 groups are used as the training set and the rest as the test set.

To verify the performance of the model, two groups of comparative experiments are set up, namely ST-GCN and CNN + LSTM. The former mainly uses the skeleton feature as input, which is the mainstream model of the skeleton behavior recognition. The latter mainly uses CNN to extract image features and LSTM to extract sequential features on the basis of image features, which is the mainstream framework of video understanding.

7.1.2. Result and discussion

The experiment is carried out on the training set of the self-built dataset. The Adam optimizer with the initial learning rate of 0.01 is set to update the parameters by gradient back propagation. The batch size is set to 256 with a total of 100 training cycles. The generated training model is then applied to the validation set for testing for statistical prediction results. The comparison of evaluation indexes is shown in Table 4.

The method proposed in this paper is much better than the other two mainstream algorithms in four indexes. The reason is that in the assembly environment, the behavior of workers performing assembly work is different from that of the ordinary human body. ST-GCN only considers the input of skeleton features. Although this approach achieves good results in recognizing ordinary human body behavior, the range of human skeleton changed is limited in the assembly behavior field. In addition, other information would lead to the poor performance

Table 5
Comparison of different features.

	Accuracy	Precision	Recall	F1-Measure
Ske&Seq	0.65	0.69	0.56	0.57
Ske&P	0.76	0.83	0.77	0.77
Seq&P	0.85	0.88	0.82	0.85
Ske&Seq&P	0.91	0.93	0.89	0.90

of the model whose accuracy is only 68%. However, the network structure of CNN+LSTM pays too much attention to the image itself. For a single assembly scene, the model directly trained can not distinguish the current assembly behavior well whose accuracy is 77%. The model proposed in this paper takes into account the characteristics of assembly behavior, integrating the human skeleton and parts features to improve the performance of the model. The proposed method achieves 91% in accuracy. It is proved that self-attention mechanism has certain advantages in human operational behavior recognition. However, the accuracy of 91% is far from meeting the requirements of industrial applications. Therefore, to improve the quality of the dataset, at the same time, the role of the self-attentive mechanism in the recognition process is deeply plumbed based on the interpretability requirements, in order to seek further methods to improve the recognition accuracy.

7.1.3. Ablation

Besides, in order to verify the impact level of the multiple features fused in the model on the final performance of the model, three sets of comparison experiments were set up in this paper, skeletal features + sequential features (Ske&Seq), skeletal features + part features

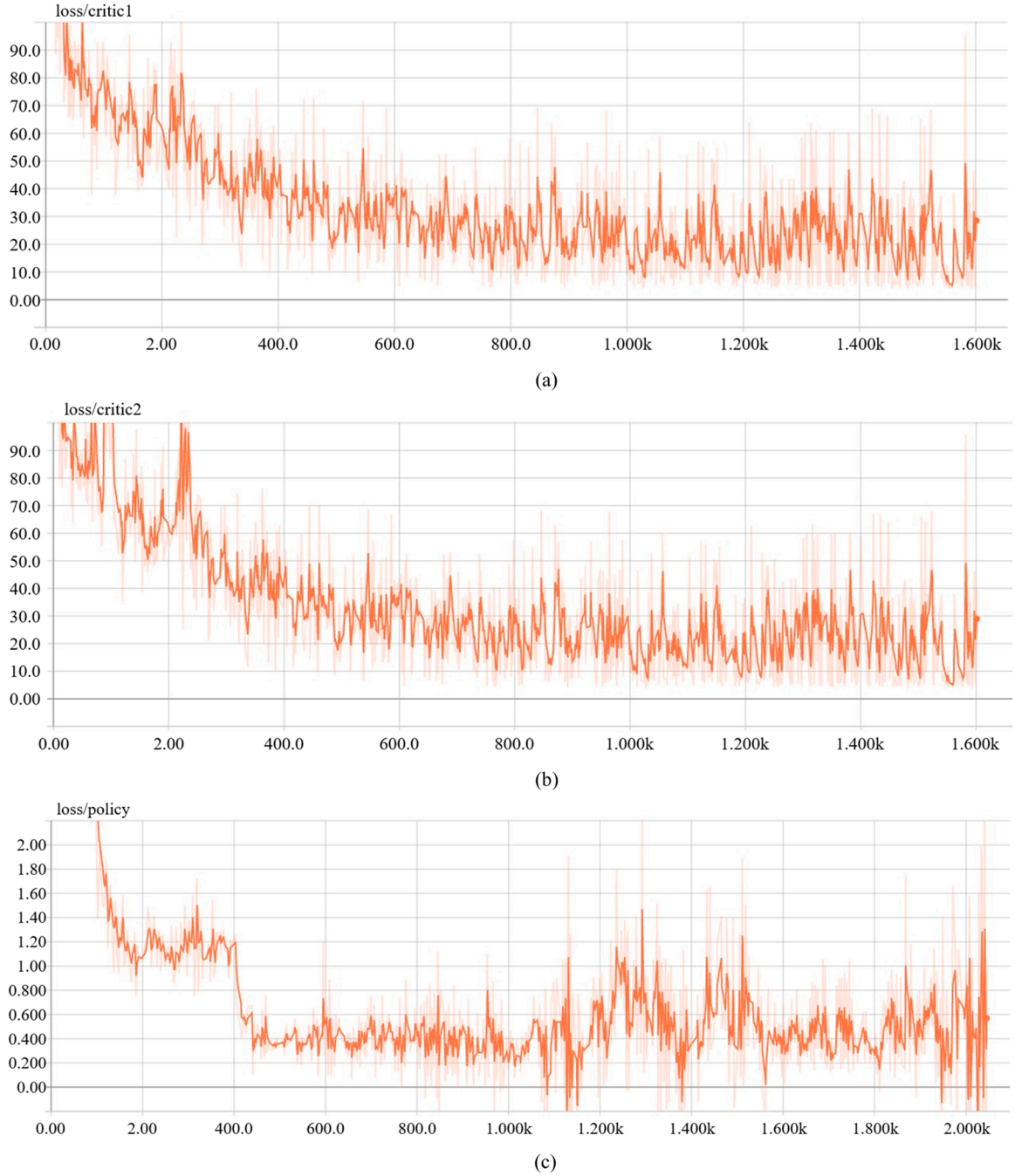


Fig. 12. The performance curves for the SAC in robot adaptive control.

(Ske&P), sequential features + part features (Seq&P). The training parameters were kept constant, and the final results of the comparison experiments obtained are shown in Table 5.

It can be seen that there is a significant decrease in the accuracy of the model after subtracting the part feature. It is inferred that the part features play a crucial role in the assembly behavior classification problem, bringing a 26% accuracy improvement to the model. On the one hand, this is because the part feature branch introduces additional training parameters to increase the capacity of the model. On the other hand, in the assembly environment, the state of the part feature largely determines the current assembly behavior. The deletion of the sequential features leads to a 15% decrease in accuracy. This result indicates that the sequential features would have a larger improvement on the performance of the model. The impact of skeletal features is only 6%, since some of the skeletal features are included in the sequential feature extraction. However, the sequential skeletal data cannot fully retain the original skeletal spatial features after the neural network, additional spatial features extracted from the original skeletal features can enhance the model.

7.2. Robot adaptive verification based on SAC

In order to verify the usability of the SAC method in the human-robot coexisting environment, a virtual robot collaboration environment in V-rep is developed with a defined self-learning grasping task for the robot. In this process, an obstacle similar to the size of the human arm to move randomly in space is set. In the training phase, the robot first searches for the required action through random policy search. The corresponding action policy is stored and then be rewarded in the experience replay pool. In the test phase, the robot selects the positive reward policy in the experience replay pool by identifying the state of the environment, and finally completes the grasping task in the dynamic environment.

The reinforcement learning algorithm based on SAC basically starts to converge at 600 steps. Fig. 12 (a) and 12 (b) show that the two critical mechanisms in the network present the same convergence state. As shown in Fig. 12 (c), although the policy network converges at step 430, there is still a large oscillation. It may due to the random dynamic obstacles actively approaching the robot and cause the interference.

8. Conclusion and future works

This study systematically introduces several key technologies involved in human-robot collaboration, the semanticized representation form of part-behavior and/or graph assembly scenarios, and demonstrates that the approach of self-attention and feature fusion facilitates the accuracy of human behavior recognition, and that reinforcement learning algorithms can help robots make adaptive decisions, thus enhancing the adjustment capability of human-robot collaboration systems for dynamic tasks. In the next research, we will work on exploring the interpretable problem of self-attention in operational behavior recognition to find more efficient and accurate solutions.

In addition, human-robot collaborative assembly in the real manufacturing process requires very low interaction delay, and is limited by the differences between the simulation model and the real assembly environment, resulting in the complexity of the real collaborative experiment. With the extensive research of digital twin technology and the development of augmented reality technology, building a digital twin operation platform for human-robot collaboration, augmented reality for efficient interaction, and human factor analysis in the collaborative process are important means to promote human-robot collaboration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is financially supported by National Key Research and Development Plan of China (Grant 2019YFB1706300), in part by Fundamental Research Funds for the Central Universities (No. 2232019D3-32), and in part by Shanghai Sailing Program (19YF1401600).

References

[1] Gualtieri L, Rauch E, Vidoni R, Matt DT. Safety, ergonomics and efficiency in human-robot collaborative assembly: design guidelines and requirements. Procedia CIRP 2020;91:367–72. https://doi.org/10.1016/j.procir.2020.02.188.
[2] Costa Mateus JE, Aghezzaf E-H, Claeys D, Limère V, Cottyn J. Method for transition from manual assembly to Human-Robot collaborative assembly. IFAC-Pap 2018; 51:405–10. https://doi.org/10.1016/j.ifacol.2018.08.328.
[3] Raatz A, Blankemeyer S, Recker T, Pischke D, Nyhuis P. Task scheduling method for HRC workplaces based on capabilities and execution time assumptions for robots. CIRP Ann 2020;69:13–6. https://doi.org/10.1016/j.cirp.2020.04.030.
[4] Li S, Wang R, Zheng P, Wang L. Towards proactive human–robot collaboration: a foreseeable cognitive manufacturing paradigm. J Manuf Syst 2021;60:547–52. https://doi.org/10.1016/j.jmsy.2021.07.017.
[5] Costanzo M, Maria GD, Lettera G, Natale C. A multimodal approach to human safety in collaborative robotic workcells. IEEE Trans Autom Sci Eng 2021:1–15. https://doi.org/10.1109/TASE.2020.3043286.
[6] Zhang Z, Qian K, Schuller BW, Wollherr D. An online robot collision detection and identification scheme by supervised learning and Bayesian decision theory. IEEE Trans Autom Sci Eng 2020:1–13. https://doi.org/10.1109/TASE.2020.2997094.
[7] Fan J, Zheng P, Li S. Vision-based holistic scene understanding towards proactive human-robot collaboration: a survey. Robot Comput-Integr Manuf 2021;75. https://doi.org/10.1016/j.rcim.2021.102304.
[8] Zanchettin AM, Casalino A, Piroddi L, Rocco P. Prediction of human activity patterns for human–robot collaborative assembly tasks. IEEE Trans Ind Inform 2019;15:3934–42. https://doi.org/10.1109/TII.2018.2882741.
[9] Rahman SMM, Wang Y. Mutual trust-based subtask allocation for human–robot collaboration in flexible lightweight assembly in manufacturing. Mechatronics 2018;54:94–109. https://doi.org/10.1016/j.mechatronics.2018.07.007.
[10] Mello LSH, de, Sanderson AC. AND/OR graph representation of assembly plans. IEEE Trans Robot Autom 1990;6:188–99. https://doi.org/10.1109/70.54734.
[11] Johannsmeier L, Haddadin S. A hierarchical human-robot interaction-planning framework for task allocation in collaborative industrial assembly processes. IEEE Robot Autom Lett 2017;2:41–8. https://doi.org/10.1109/LRA.2016.2535907.
[12] Aliev K, Antonelli D, Bruno G. Task-based programming and sequence planning for human-robot collaborative assembly. IFAC-Pap 2019;52:1638–43. https://doi.org/10.1016/j.ifacol.2019.11.435.
[13] Nian R, Liu J, Huang B. A review on reinforcement learning: introduction and applications in industrial process control. Comput Chem Eng 2020;139:106886. https://doi.org/10.1016/j.compchemeng.2020.106886.
[14] Ishida F, Sasaki T, Sakaguchi Y, Shimai H. Reinforcement-learning agents with different temperature parameters explain the variety of human action–selection behavior in a Markov decision process task. Neurocomputing 2009;72:1979–84. https://doi.org/10.1016/j.neucom.2008.04.009.
[15] Haarnoja, T., Zhou, A., Abbeel, P., Levine, S.. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. ArXiv: 180101290 [Cs, Stat] 2018.
[16] Gualtieri L, Rauch E, Vidoni R, Matt DT. An evaluation methodology for the conversion of manual assembly systems into human-robot collaborative workcells. Procedia Manuf 2019;38:358–66. https://doi.org/10.1016/j.promfg.2020.01.046.
[17] Papanastasiou S, Kousi N, Karagiannis P, Gkournelos C, Papavasileiou A, Dimoulas K, et al. Towards seamless human robot collaboration: integrating multimodal interaction. Int J Adv Manuf Technol 2019;105:3881–97. https://doi.org/10.1007/s00170-019-03790-3.
[18] Duan F, Tan JTC, Tong JG, Kato R, Arai T. Application of the assembly skill transfer system in an actual cellular manufacturing system. IEEE Trans Autom Sci Eng 2012;9:31–41. https://doi.org/10.1109/TASE.2011.2163818.
[19] Raessa M, Chen JCY, Wan W, Harada K. Human-in-the-loop robotic manipulation planning for collaborative assembly. IEEE Trans Autom Sci Eng 2020;17:1800–13. https://doi.org/10.1109/TASE.2020.2978917.
[20] Ravichandar HC, Dani AP. Human intention inference using expectation-maximization algorithm with online model learning. IEEE Trans Autom Sci Eng 2017;14:855–68. https://doi.org/10.1109/TASE.2016.2624279.
[21] Lee D, Liu C, Liao Y, Hedrick JK. Parallel interacting multiple model-based human motion prediction for motion planning of companion robots. IEEE Trans Autom Sci Eng 2017;14:52–61. https://doi.org/10.1109/TASE.2016.2623599.
[22] Andrianakos G, Dimitropoulos N, Michalos G, Makris S. An approach for monitoring the execution of human based assembly operations using machine learning. Procedia CIRP 2019;86:198–203. https://doi.org/10.1016/j.procir.2020.01.040.

[23] Wang XV, Kemény Z, Váncza J, Wang L. Human–robot collaborative assembly in cyber-physical production: classification framework and implementation. CIRP Ann 2017;66:5–8. https://doi.org/10.1016/j.cirp.2017.04.101.
[24] Wang Q, Jiao W, Yu R, Johnson MT, Zhang Y. Virtual reality robot-assisted welding based on human intention recognition. IEEE Trans Autom Sci Eng 2020;17:799–808. https://doi.org/10.1109/TASE.2019.2945607.
[25] Liu H, Wang L. Collision-free human-robot collaboration based on context awareness. Robot Comput-Integr Manuf 2021;67:101997. https://doi.org/10.1016/j.rcim.2020.101997.
[26] Ding H, Wan Z, Zhou Y, Tang J. A data-driven programming of the human-computer interactions for modeling a collaborative manufacturing system of hypoid gears by considering both geometric and physical performances. Robot Comput-Integr Manuf 2018;51:121–38. https://doi.org/10.1016/j.rcim.2017.10.003.
[27] Lin C-H, Wang K-J, Tadesse AA, Woldegiorgis BH. Human-robot collaboration empowered by hidden semi-Markov model for operator behaviour prediction in a smart assembly system. J Manuf Syst 2022;62:317–33. https://doi.org/10.1016/j.jmsy.2021.12.001.
[28] Mei K, Zhang J, Li G, Xi B, Zheng N, Fan J. Training more discriminative multi-class classifiers for hand detection. Pattern Recognit 2015;48:785–97. https://doi.org/10.1016/j.patcog.2014.09.001.
[29] Nan Z, Shu T, Gong R, Wang S, Wei P, Zhu S-C, et al. Learning to infer human attention in daily activities. Pattern Recognit 2020;103:107314. https://doi.org/10.1016/j.patcog.2020.107314.
[30] Liu Z, Liu Q, Xu W, Liu Z, Zhou Z, Chen J. Deep learning-based human motion prediction considering context awareness for human-robot collaboration in manufacturing. Procedia CIRP 2019;83:272–8. https://doi.org/10.1016/j.procir.2019.04.080.
[31] Faber M, Mertens A, Schlick CM. Cognition-enhanced assembly sequence planning for ergonomic and productive human–robot collaboration in self-optimizing assembly cells. Prod Eng Res Dev 2017;11:145–54. https://doi.org/10.1007/s11740-017-0732-9.
[32] Bannat A, Bautze T, Beetz M, Blume J, Diepold K, Ertelt C, et al. Artificial cognition in production systems. IEEE Trans Autom Sci Eng 2011;8:148–74. https://doi.org/10.1109/TASE.2010.2053534.
[33] Liu H, Wang L. Human motion prediction for human-robot collaboration. J Manuf Syst 2017;44:287–94. https://doi.org/10.1016/j.jmsy.2017.04.009.
[34] Sadrfaridpour B, Wang Y. Collaborative assembly in hybrid manufacturing cells: an integrated framework for human–robot interaction. IEEE Trans Autom Sci Eng 2018;15:1178–92. https://doi.org/10.1109/TASE.2017.2748386.
[35] Tsarouchi P, Matthaiakis A-S, Makris S, Chryssolouris G. On a human-robot collaboration in an assembly cell. Int J Comput Integr Manuf 2017;30:580–9. https://doi.org/10.1080/0951192X.2016.1187297.
[36] Arana-Arexolaleiba N, Urrestilla-Anguiozar N, Chrysostomou D, Bøgh S. Transferring human manipulation knowledge to industrial robots using reinforcement learning. Procedia Manuf 2019;38:1508–15. https://doi.org/10.1016/j.promfg.2020.01.136.
[37] Sun X, Zhang R, Liu S, Lv Q, Bao J, Li J. A digital twin-driven human–robot collaborative assembly-commissioning method for complex products. Int J Adv Manuf Technol 2022;118:3389–402. https://doi.org/10.1007/s00170-021-08211-y.
[38] Li X, Zhong J, Kamruzzaman MM. Complicated robot activity recognition by quality-aware deep reinforcement learning. Future Gener Comput Syst 2021;117:480–5. https://doi.org/10.1016/j.future.2020.11.017.
[39] Loftus TJ, Filiberto AC, Li Y, Balch J, Cook AC, Tighe PJ, et al. Decision analysis and reinforcement learning in surgical decision-making. Surgery 2020;168:253–66. https://doi.org/10.1016/j.surg.2020.04.049.
[40] Huang K, Ma X, Song R, Rong X, Tian X, Li Y. A self-organizing developmental cognitive architecture with interactive reinforcement learning. Neurocomputing 2020;377:269–85. https://doi.org/10.1016/j.neucom.2019.07.109.
[41] Knox WB, Stone P. Framing reinforcement learning from human reward: reward positivity, temporal discounting, episodicity, and performance. Artif Intell 2015;225:24–50. https://doi.org/10.1016/j.artint.2015.03.009.
[42] Oliff H, Liu Y, Kumar M, Williams M, Ryan M. Reinforcement learning for facilitating human-robot-interaction in manufacturing. J Manuf Syst 2020;56:326–40. https://doi.org/10.1016/j.jmsy.2020.06.018.