Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 121 (2024) 1–6

www.elsevier.com/locate/procedia

11th CIRP Global Web Conference (CIRPe 2023)

Selective disassembly planning considering process capability and component quality utilizing reinforcement learning

Roham Sadeghi Tabar[a,*], Maria Chiara Magnanini[b], Florian Stamer[c], Marvin Carl May[c], Gisela Lanza[c], Kristina Wärmefjord[a], Rikard Söderberg[a]

[a] *Chalmers University of Technology, SE-41296 Gothenburg , Sweden*
[b] *Politecnico di Milano, 20133 Milan , Italy*
[c] *wbk Institute of Production Science, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany*

* Roham Sadeghi Tabar. Tel.: +46-031-772-6745. *E-mail address:* rohams@chalmers.se

Abstract

Disassembly is a crucial process for achieving circular products, enabling function recovery, material reuse, and recycling. Disassembly planning is complex due to epistemic uncertainty associated with each unique product's conditions, i.e., quality and aleatoric uncertainty about the capabilities of available resources and processes, and the cost benefits of associated operations impede planning. Therefore, the disassembly is intended to result in keeping the maximum value for the disassembled units of the product. In selective disassembly, the specification of the units of the product to be disassembled is acquired, leaving the rest of the product intact. The benefit of selective disassembly is to minimize waste during dismantling and maximize the reuse of the disassembled components for economic and ecological sustainability. The challenges in disassembly sequence planning include product complexity, operational and technological process capabilities, and the lack of information regarding the product architecture. For this complex planning task, limited studies have been performed on incorporating process capabilities with respect to the operations resources for selective disassembly planning. In this paper, an approach for optimal sequence planning of the selective disassembly process is put forward, taking into account multiple constraints, i.e., quality, time, and process capability. The intelligent planning approach takes advantage of a reinforcement learning model to handle the complexity of the planning problem. The approach has been implemented and tested on an industrial reference assembly. The result shows that the complex task of selective disassembly planning can be efficiently performed utilizing the proposed approach.

© 2023 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 11th CIRP Global Web Conference

Keywords: Remanufacturing; Selective Disassembly; Planning; Reinforcement Learning

1. Introduction

Given the global challenges of reducing resource consumption and waste, reusing products in closed-loop systems offers a solution. One enabler for achieving a closed-loop system has been remanufacturing. Remanufacturing extends the lifespan of products and minimizes waste while decreasing the environmental impact [1]. Product disassembly has been an elementary aspect in the remanufacturing strategies enabling the separation of components for reuse, recovery, and recycling [6]. However, the high manual labor involved in remanufacturing reduces its economic feasibility.

Automation can make remanufacturing more cost-effective, but fully automated disassembly processes are not yet established in the industry. Hybrid production systems combining manual and automated workstations show a promising tradeoff due to the flexibility required for handling product variety [18]. While there have been attempts to automate disassembly for various products, only a few applications exist on an industrial scale, such as Apple's robot-based disassembly line for smartphones.

Advances in artificial intelligence (AI) and robotics offer the potential for flexible and autonomous production systems. Machine learning, particularly deep reinforcement learning, allows robots to adapt and solve problems similar to humans. Autonomous disassembly robots can improve productivity and tackle tasks with high uncertainty.

The first challenge for such a system is the identification of the reusable and repairable material during the disassembly process imposing uncertainty in the planning strategies. To address these challenges, previous research has focused on introducing

2212-8271 © 2023 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 11th CIRP Global Web Conference
10.1016/j.procir.2023.09.221

planning perspectives, where only the required components are separated and reused from the end-of-life (EoL) products. This approach has been referred to as the selective disassembly [15]. This paper focuses on the selective disassembly problem utilizing AI, considering process capability and component quality. Concretely, the application is based on using Reinforcement Learning (RL), which is a powerful production control technique[8].

In the next section, the overview of the research performed in this field is presented.

1.1. Disassembly Planning

Disassembly planning involves the systematic determination of the sequence of operations for disassembling complex objects into their constituent components. One major challenge in the planning and control of this process is the broad spectrum of EoL products. This entails the variation of the discarded products, the uncertain properties, and the assembly structure [9, 20]. Recent studies have addressed these challenges by introducing planning strategies for the disassembly process. Sontochi et al. have introduced a computer-aided disassembly planning approach to help the designer and disassembler in the optimization formulation for design for recycling. In this approach, the product analysis, operation sequence generation, disassembly tool selection, and economic evaluation are performed based on the CAD, process, and lifecycle databases [14]. In a recent review, Chang et al. have summarized the challenges with disassembly planning on a detail level, where the geometry of each part and component is analyzed, sequence level and task planning, and reveres logistics [4]. The utilized methods for optimization perspectives for planning have mainly been based on graph-based and geometry-based heuristics, metaheuristics, i.e., evolutionary algorithms such as GA and particle swarm. Similarly, for assembly process sequence planning, heuristics and evolutionary algorithms have been utilized extensively [13, 12]. As Santochi et al. has suggested [14], many of the methods utilize CAD structure to create the precedence matrices and thereby formulate the binary optimization algorithms. With the growing state space of the complex products, the traditional decomposition of the product to all the components has been considered redundant. For complex structures, where there are many interacting surfaces between the components, the number of alternatives for disassembly increases, and since the problem is binary, the combinatorial explosion occurs, thus the heuristics are often utilized to solve the selective disassembly problem. Thus, planning strategies have been proposed to recover subassemblies of components when needed, i.e., selective disassembly. In the next section, the details of this approach are introduced.

1.2. Selective disassembly

Selective disassembly targets one or several components of a product and determines the sequence of disassembly operations to extract them [7]. Kara et al. have pinpointed that the complete disassembly of a product is unproductive due to cost constraints and product conditions and thereby suggested a selective disassembly approach with a limited number of disassembly paths based on liaisons diagrams for sequence generation and rule-based constraints on disassembly sequences [6]. One challenge in the disassembly process is the uncertainty in the operation time for each disassembly step. To address this challenge, a selective disassembly formulation with random operation times in the parallel disassembly environments is considered, where an integer programming approach is proposed to solve the optimal sequence generation problem [7]. To improve solution quality, model complexity, and search time, a graph-based approach is combined with expert rules to efficiently search for the optimal operation sequences [15]. Several studies have considered combined graph methods for sequence representation and meta-heuristic algorithms for solving the optimal sequence generation. Utilizing this approach, the Artificial Bee Colony meta-heuristic is combined with an AND/OR graph representation [17]. Similarly AND/OR Graphs and Genetic algorithms have been implemented to smoothen the search process during the optimization [10]. In a recent review of the selective disassembly approaches, machine learning approaches have been specified as an area to be explored for planning strategies [3]. Reinforcement learning has been previously applied to the selective disassembly planning problem. Deep Q-networks have been applied to cope with the product structure and operation uncertainty [22, 19]. Several Q- learning approaches have been introduced for disassembly planning for maximizing component value after disassembly [11, 2]. In a recent review of the reinforcement learning-based disassembly planning approaches, it is identified that 50% of the publications (out of 16 included in the comparison) have a time objective. Furthermore, 38% of the publications have a cost perspective as the main objective [5]. Although many strategies for representation and optimization approaches have been proposed for selective disassembly, less research has been performed on combining the process capability of the disassembly process into time variation [21]. Quality perspectives of the extracted material have also been majorly neglected in the selected criteria for disassembly. In this paper, these two perspectives have been included in the formulation of the selective disassembly problem, and a reinforcement learning approach is proposed for solving optimal sequence generation.

1.3. Scope of the paper

Several studies have been performed on the representation and optimization perspective for the disassembly process. While most of the research has been on operation time for each disassembly task, the technological capabilities have been neglected, which are of utmost importance when dealing with circular economy. The geometric quality of the extracted components has also been disregarded likewise. In this paper, a reinforcement learning approach has been considered for solving the optimal sequence generation for the disassembly process, while the quality and process capability have been integrated into the optimization criteria. The first Section provided a background to the problem and a literature review of the available

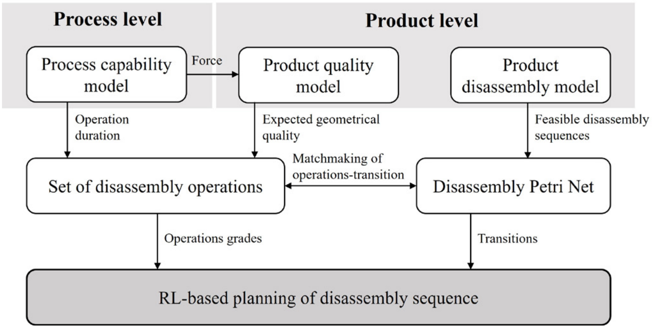

Fig. 1. Schematic overview of the proposed methodology

methods. The rest of the paper is structured as follows. Section 2 provides the problem formulation and the proposed method. Section 3 presents the selected reference case for this study and describes the implementation of the approach. Section 4 presents the results and discussions of these and is followed by a conclusion in Section 5.

2. Method

In this section, the modeling details of the selective disassembly approach are presented.

The schematic representation of the proposed methodology is presented in Fig.1. The selected disassembly approach grounds on the integration of process and product information at several levels.

This approach could be further generalized to include additional process capability information, such as technological details, and additional product information, as expected dimensional deviation etc. In the following, the disassembly model based on process capability characteristics and product quality features is described. Then, the formulation of the objective is presented, and later the details of the reinforcement learning approach are put forward.

2.1. Disassembly model

To model the disassembly process, we propose a Markov decision process (MDP) framework. The MDP will be used as a learning environment for reinforcement learning. For this, a 5-tuple MDP is established and is represented by $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$ where $\mathcal{S}$ represents the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the probability of state transition, and $\mathcal{R}$ represents the reward of taking each transition [8]. In the context of selective disassembly, the transition is performed to inquire about the specific component. In order to achieve the desired component, the directed graph of the disassembly process is established as a result of an initial Disassembly Petri Net (DPN). Utilizing a DPN, we describe the product matching conditions and precedence relationship. The DPN is formulated with five tuples as $(\mathcal{Z}, \mathcal{T}, \mathcal{I}, \mathcal{O}, \mathcal{M}_0)$. In this DPN, $\mathcal{Z}$ is a set of places similar to states. $\mathcal{T}$ are the transitions that can be interpreted as actions in the MDP. $\mathcal{I}$ is a $n \times m$ matrix, where n is the number of places and m denotes the number of transitions. This matrix represents the place-transition input arcs. Matrix $\mathcal{O}$ is the transition-places output arcs. $\mathcal{M}_0$ is the initial state of the disassembly, i.e., the EoL product. With this formulation, a new state for the disassembly can be calculated as:

$$\mathcal{M} = \mathcal{M}_0 + \mathcal{T}_F \cdot (\mathcal{O} - \mathcal{I}^T), \quad (1)$$

here, $\mathcal{T}_F$ is a binary vector stating fired transitions.

We consider the objective of selective disassembly as the accumulative value of the quality and process capability indexes. For the process capability index $\mathcal{P}$, we consider three grades (G), the operations that have short(Grade 3), medium (Grade 2), and long task (Grade 1) times. These three grades categorize the disassembly operations for manual and automatic processes.

Table 1. Operation grades according to the process capabilities and force (G =$\{1, 2, 3\}$)

Operation	Automatic Capability (G)	Force (G)	Manual Process Capability(G)	Force(G)
Fixturing	1	2	3	1
Gripping	3	2	3	1
Unscrewing	2	3	1	3
Pulling	1	1	1	2

These grades are considered as the reward of a transition in the MDP, for each disassembly transition.

The disassembly forces with respect to each operation are also graded as low (Grade 3), medium (Grade 2) and large forces (Grade 1). For the geometric quality criteria, we consider the surface profile of the components after disassembly $\mathcal{D}$. We assume a linear relationship for the force-deviation and establish the following:

$$\mathcal{K}_T = \mathcal{F} \times \mathcal{D} \quad (2)$$

Thereby, the tolerance zone of each component is divided into three zones, low, medium, and high. For each component, the deviation and forces are registered and $\mathcal{K}_T$ is calculated, and the surface profile grade is identified and graded as G=3 for low values, two for medium, and one for high, respectively. Furthermore, the state of wear of the product's functional surfaces is identified and graded as an indication of the recovery value of the component. Similarly, a low-medium-high grade is considered for the component's wear and is assigned as a quality measure of the extracted component. We denote the wear value as $\mathcal{K}_W$. We consider the objective of the disassembly as the problem of maximizing the reward during the actions of the MDP to retrieve the desired component.

2.2. Sequence planning with reinforcement learning

To maximize the reward, which is the objective of the planning, a Q-learning approach is utilized for training an agent to make sequential decisions on which components or parts to disassemble. The disassembly model specified in Section 2.1 is utilized as the environment, providing the state-space specification for the agent to take action. We consider a discrete observation and action space with the MDP. The general cycle of

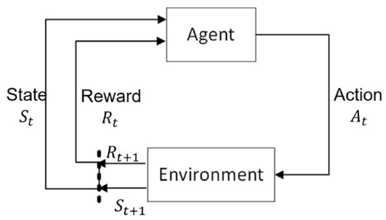

Fig. 2. Schematic reinforcement learning cycle

reinforcement learning (RL) is visualized in Fig. 2. Note that t and $t+1$ refer to the current and next-time steps. The agents perception of the current state of the environment is called state S_t, which is used by the agent to evaluate a suitable action A_t to be selected at time-step t in order to change the environment. After executing A_t, the environment transfers to S_{t+1}. As shown in Fig. 2, this cycle continues while after each action A, the agent receives a reward signal R_t, which is used to improve the agent's mapping from state S to action A [16]. The Q-learning agent here is a value-based RL agent that trains a critic to estimate the accumulated rewards. The critic is a value function approximator. We refer to the critic as $Q(S, A, \phi)$, where ϕ is the approximation parameter.The expected long-term reward for the S and A as inputs are approximated by the critic, and the Q values are stored in a Q-value table. The training algorithm of the Q-learning agent with epsilon-greedy exploration is composed of the following until the terminal state S is achieved:

- Initialization with the observation S.
- Select a random action A with ϵ probability.

$$A = arg\ max_A Q(S, A; \phi) \tag{3}$$

- Approximate the reward and calculate the next state S_{t+1}.
- Set the Q-value to y with the discount factor γ:

$$y = R + \gamma\ max_A\ Q(S_{t+1}, A, \phi) \tag{4}$$

  The discount factor determines the impact of the reward on the choice of the next action.
- Calculate the Q-value difference ΔQ .
- Update the critic with Q-value with the learning rate α.

$$Q(S, A) = Q(S, A, \phi) + \alpha.\Delta Q \tag{5}$$

For the reward function, we introduce the aggregated reward of the process capability and the quality criteria as introduced in the previous section. Process capability reward for each operation, denoted by $\mathcal{P}^i$, surface quality reward denoted by $\mathcal{K}_T^i$ and the state of wear of the components after disassembly denoted by $\mathcal{K}_W^i$. We calculate the reward as:

$$R^i = \sum_{i=1}^{m} z_1.\mathcal{P}^i + z_2.\mathcal{K}_T^i + z_3.\mathcal{K}_W^i \tag{6}$$

Here, z_1 to z_3 are the weights for each of the indexes.

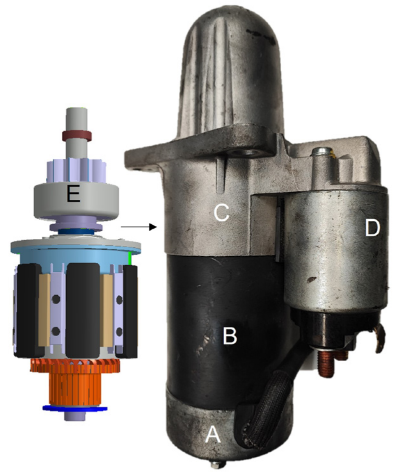

Fig. 3. Starter motor engine and its components

In the next section, the proposed assembly modeling and selective assembly sequence generation is applied to the reference case.

3. Reference case and implementation

To evaluate the proposed selective disassembly approach, a reference case has been chosen, and the disassembly model is established. The Q-learning algorithm is applied to the case, and the disassembly sequence is generated. The following presents the details of the reference case and the implementation of the approach.

3.1. Case description

The case focuses on an engine starter motor. Engine starter motors are a great example, as they are already remanufactured in the industry. However, the disassembly is purely manual and relies on the expertise of the worker. The starter motor consists of several critical components. The first component is the solenoid which brings the starter in and out of the mechanical chain of the engine, part D in Fig. 3. The second critical part is the gear, part E. In the gear, teeth can be broken, and thus the function can be negatively affected. Another critical part is the commutator. As the motor is brushed, the carbon brushes are slowly worn out, Fig. 3, Part E. For simplification, the starter motor is considered to be composed of 5 main components; The casing lock, part A, the engine housing, B, the gear housing, C, the solenoid, D, and the inner engine components, part E. The goal of the selective disassembly is to recover the engine's inner components, part E.

3.2. Implementation

To establish the environment for the QL algorithm, first, the DPN is constructed. The DPN is visualized in Fig. 4, where the state space is visualized with the circles, and the arcs represent

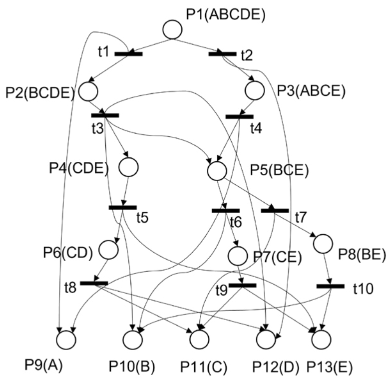

Fig. 4. Starter motor disassembly Petri Net

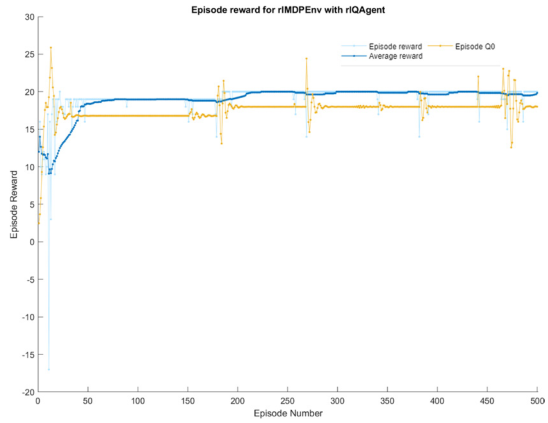

Fig. 5. Training episode of the QL for the reference case with automatic process

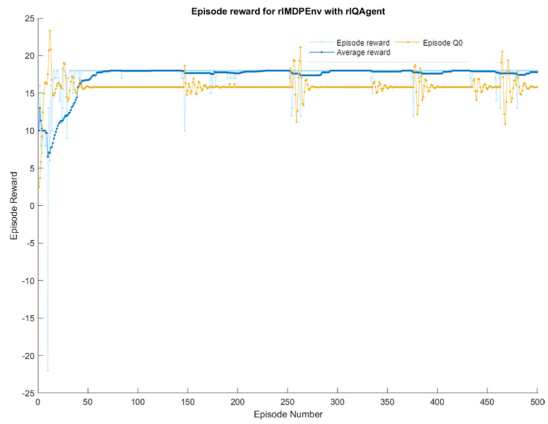

Fig. 6. Training episode of the QL for the reference case with manual process

the transitions, i.e., disassembly operations. The states, P_6, P_9 to P_{13}, are defined as terminal states of the MDP where component E is extracted. The MDP environment and the QL algorithm are set up with MATLAB. The learning rate α has been set to one. Similarly, the ϵ and the decay parameters have been set to 0.9 and 0.01, respectively. The agent has been trained with 500 episodes, where the maximum step per episode has been 50. The reward function is established according to equation 6 and is connected to the environment to approximate the Q-values. The agent is trained on a workstation with seven cores 2.7 GHz processor, and 32 GB of RAM.

Two scenarios have been considered in setting up the reward function. One performs QL, where the rewards are based on manual operations, and one for automatic operations, according to the description in section 2. The initial state for the started motor has been a complete product, with all the components A to E, and the sequence is terminated when the component E is extracted.

The next section presents the results of applying the proposed selective disassembly method and the QL learning to identify the optimal sequence of disassembly.

Table 2. Optimal disassembly sequence

Operation	Sequence	Average Reward
Automatic	ABCDE-BCDE-BCE-CE-E	19
Manual	ABCDE-ABCE-BCE-CE-E	18

4. Discussion

The QL algorithm is trained on the defined MDP based on the DPN description in Fig. 4. The episode reward evolution for the automatic and the manual processes are depicted in Fig. 5 and Fig. 6. The optimal sequences achieved for each process are reported in Table 2. Two distinct sequences are achieved for each process. For the automatic process, the optimal sequence is by removing the engine casing lock, followed by removing the solenoid and the engine casing. Finally, the engine component E is extracted from the assembly, and the objective is achieved with the reward of 19. For the manual process, the optimal achieved sequence is removing the solenoid first, unscrewing the top lock, gripping and pulling part B and finally removing engine part E. The average achieved reward for the manual process has been 18. The main differences in the achieved optimal sequence are due to different process capabilities of the manual and automatic processes, where the fixturing and centering of the components during disassembly and the gripping processes are rewarded with different grades, based on the division established in Table 1. Moreover, different force grades achieved by different tasks during the disassembly have caused the accumulated reward to differ based on the defined quality criteria. This also pinpoints that integration of the quality of the components after each disassembly operation results in changes in the disassembly plan. Another remarkable point is the fast ramp-up of the learning process for both of the agents. This can be argued from the perspective of the limited state space due to the size of the problem. It is expected that with increasing the state space and complexity of the product and, thereby, the environment, the training time increase, and the learning curve dampens.

Furthermore, in this study, we have considered a known product structure by which the DPN could be generated and QL trained. For uncertain product structures, ideally, a batch of products of the same variant should be available for the RL to be trained.

5. Conclusion

The problem of optimal sequence generation for selective disassembly is studied. This paper has proposed a reinforcement learning approach with QL, which is established by a disassembly model based on a disassembly Petri net. For the integration of the process capability and the quality perspective into the decision making, an aggregated reward function of the criteria has been proposed. The distinction of reward grades has been established based on the specific disassembly operations. The proposed approach has been applied to a starter engine motor with the objective of maximizing the reward of the disassembly sequence with respect to the process capability and quality. The goal of the disassembly has been to retrieve the inner engine components. The trained agents have provided two distinct sequences considering automatic and manual disassembly processes. Based on the results achieved, it is concluded that integration of the quality and process capability perspectives into the learning algorithms impact the decision of the optimal sequences and, thereby need to be considered as a decision criterion.

Future research includes exploring the behavior of the designed algorithm on problems with large state space and additional physics-based characterization of process and product features. In this case, a disassembly simulation environment is to be integrated into the QL, where the agent can directly interact with the geometries. Moreover, the authors intend to enhance the reward function with real-time feedback from a simulation environment with respect to the designed algorithm for enhanced accuracy of representation.

Acknowledgments

This work was carried out at the Wingquist Laboratory within the Area of Advance Production at Chalmers, in collaboration with wbk Institute of Production Science at Karlsruhe Institute of Technology (KIT) and supported by the Swedish Innovation Agency (VINNOVA). The support is gratefully acknowledged.

References

[1] Atasu, A., Sarvary, M., Van Wassenhove, L.N., 2008. Remanufacturing as a marketing strategy. Management science 54, 1731–1746.
[2] Bi, Z., Guo, X., Wang, J., Qin, S., Qi, L., Zhao, J., 2022. A q-learning-based selective disassembly sequence planning method, in: 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3216–3221. doi:10.1109/SMC53654.2022.9945073.
[3] Chand, M., Ravi, C., 2023. A state-of-the-art literature survey on artificial intelligence techniques for disassembly sequence planning. CIRP Journal of Manufacturing Science and Technology 41, 292–310.
[4] Chang, M., Ong, S., Nee, A., 2017. Approaches and challenges in product disassembly planning for sustainability. Procedia Cirp 60, 506–511.
[5] Guo, X., Bi, Z., Wang, J., Qin, S., Liu, S., Qi, L., 2023. Reinforcement learning for disassembly system optimization problems: A survey. International Journal of Network Dynamics and Intelligence , 1–14.
[6] Kara, S., Pornprasitpol, P., Kaebernick, H., 2006. Selective disassembly sequencing: a methodology for the disassembly of end-of-life products. CIRP annals 55, 37–40.
[7] Kim, H.W., Park, C., Lee, D.H., 2018. Selective disassembly sequencing with random operation times in parallel disassembly environment. International Journal of Production Research 56, 7243–7257.
[8] Kuhnle, A., May, M.C., Schäfer, L., Lanza, G., 2022. Explainable reinforcement learning in production control of job shop manufacturing system. International Journal of Production Research 60, 5812–5834. doi:10.1080/00207543.2021.1972179.
[9] Lee, D., Kang, J., Xirouchakis, P., 2001. Disassembly planning and scheduling: review and further research. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 215, 695–709.
[10] Parsa, S., Saadat, M., 2019. Intelligent selective disassembly planning based on disassemblability characteristics of product components. The International Journal of Advanced Manufacturing Technology 104, 1769–1783.
[11] Reveliotis, S., 2004. Modelling and controlling uncertainty in optimal disassembly planning through reinforcement learning, in: IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004, pp. 2625–2632 Vol.3. doi:10.1109/ROBOT.2004.1307457.
[12] Sadeghi Tabar, R., Wärmefjord, K., Söderberg, R., 2021. Rapid sequence optimization of spot welds for improved geometrical quality using a novel stepwise algorithm. Engineering Optimization 53, 867–884.
[13] Sadeghi Tabar, R., Wärmefjord, K., Söderberg, R., Lindkvist, L., 2019. A novel rule-based method for individualized spot welding sequence optimization with respect to geometrical quality. Journal of Manufacturing Science and Engineering 141, 111013.
[14] Santochi, M., Dini, G., Failli, F., 2002. Computer aided disassembly planning: state of the art and perspectives. CIRP Annals 51, 507–529.
[15] Smith, S.S., Chen, W.H., 2011. Rule-based recursive selective disassembly sequence planning for green design. Advanced Engineering Informatics 25, 77–87.
[16] Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.
[17] Tian, G., Ren, Y., Feng, Y., Zhou, M., Zhang, H., Tan, J., 2018. Modeling and planning for dual-objective selective disassembly using and/or graph and discrete artificial bee colony. IEEE Transactions on Industrial Informatics 15, 2456–2468.
[18] Tolio, T., Bernard, A., Colledani, M., Kara, S., Seliger, G., Duflou, J., Battaia, O., Takata, S., 2017. Design, management and control of demanufacturing and remanufacturing systems. CIRP Annals 66, 585–609.
[19] Tuncel, E., Zeid, A., Kamarthi, S., 2014. Solving large scale disassembly line balancing problem with uncertainty using reinforcement learning. Journal of Intelligent Manufacturing 25, 647–659.
[20] Wiendahl, H.P., Seliger, G., Perlewitz, H., Bürkner, S., 1999. A general approach to disassembly planning and control. Production Planning & Control 10, 718–726.
[21] Wurster, M., Michel, M., May, M.C., Kuhnle, A., Stricker, N., Lanza, G., 2022. Modelling and condition-based control of a flexible and hybrid disassembly system with manual and autonomous workstations using reinforcement learning. Journal of Intelligent Manufacturing 33, 575–591. doi:10.1007/s10845-021-01863-3.
[22] Zhao, X., Li, C., Tang, Y., Cui, J., 2021. Reinforcement learning-based selective disassembly sequence planning for the end-of-life products with structure uncertainty. IEEE Robotics and Automation Letters 6, 7807–7814.