



강화학습 논문 리뷰 스터디 9기

# A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling

김용회(Kim Yong Hae)

# Agenda

---

- 논문 내용 정리
-

---

# 논문 내용 정리

---

COMPLEX SYSTEM MODELING AND SIMULATION  
ISSN 2096-9929 01/06 pp 257-270  
Volume 1, Number 4, December 2021  
DOI: 10.23919/CSSMS.2021.0027

## A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling

Ling Wang\*, Zixiao Pan, and Jingjing Wang

**Abstract:** As the critical component of manufacturing systems, production scheduling aims to optimize objectives in terms of profit, efficiency, and energy consumption by reasonably determining the main factors including processing path, machine assignment, execute time and so on. Due to the large scale and strongly coupled constraints nature, as well as the real-time solving requirement in certain scenarios, it faces great challenges in solving the manufacturing scheduling problems. With the development of machine learning, Reinforcement Learning (RL) has made breakthroughs in a variety of decision-making problems. For manufacturing scheduling problems, in this paper we summarize the designs of state and action, tease out RL-based algorithm for scheduling, review the applications of RL for different types of scheduling problems, and then discuss the fusion modes of reinforcement learning and meta-heuristics. Finally, we analyze the existing problems in current research, and point out the future research direction and significant contents to promote the research and applications of RL-based scheduling optimization.

**Key words:** Reinforcement Learning (RL); manufacturing scheduling; scheduling optimization

### 1 Introduction

Production scheduling is a crucial connecting component in the manufacturing system. To improve the production efficiency and effectiveness, scheduling algorithms play an important role, which have always been a significant research topic in interdisciplinary fields, like industrial engineering, automation, management science, and so on. Production scheduling algorithms mainly include three categories, accurate algorithms, heuristics, and meta-heuristics. The exact algorithm can guarantee to obtain the optimal solution in theory, but it is difficult to solve the large-scale problems efficiently and effectively due to the NP-hard nature. Heuristics adopt some rules to construct scheduling solutions efficiently but without global optimization perspective. Moreover, the design of rules

highly depends on the deep understanding of the problem specific characteristics. Meta-heuristics can obtain excellent scheduling solutions within acceptable computation time, but the design of search operators is seriously problem dependent. At the same time, for large-scale problems the iterative search process is very time-consuming and difficult to be applied for real-time scenarios, such as Meituan on-line food delivery.

With the development of artificial intelligence, Reinforcement Learning (RL) has been successfully applied to the sequential decision-making problems, such as games<sup>[1]</sup> and robots control<sup>[2]</sup>. During recent years, RL has been successfully applied to solve several combinatorial optimization problems, such as Vehicle Routing Problem<sup>[3]</sup> (VRP) and Traveling Salesman Problem<sup>[4]</sup> (TSP). Supposing a production scheduling problem can be regarded as the environment of RL, an agent can learn a policy or rule via reasonable designs of actions and states, as well as interaction with the environment through a large number of offline training. Such a new idea provides a novel approach for solving scheduling problems, especially the uncertain and dynamic problems with

\* Ling Wang, Zixiao Pan, and Jingjing Wang are with the Department of Automation, Tsinghua University, Beijing 100084, China. E-mail: wangling@tsinghua.edu.cn; pzx19@mails.tsinghua.edu.cn; wjj18@mails.tsinghua.edu.cn.

\* To whom correspondence should be addressed.  
Manuscript received: 2021-10-21; accepted: 2021-11-22

© The author(s) 2021. The articles published in this open access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

## ■ A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling

■ 2021. 12

### ■ Review Paper

### ■ 주요 내용 정리

- Manufacturing Scheduling은 제조 시스템 핵심 요소로서 공정 경로, 할당, 실행 시간 등 주요 요소를 합리적으로 결정하여 이익, 효율 등 설정된 목표를 최적화 하는 것을 목표로 함
- 대규모 생산 환경과 복잡한 제약 조건들의 특성 등을 실시간으로 스케줄링 하는 것은 현실적으로 매우 어려운 문제임
- ML/DL 기술의 발전과 함께 RL은 다양한 의사 결정 문제에서 획기적인 발전을 이루고 있음
- 스케줄링을 위한 RL 기반의 알고리즘을 설명
- RL과 메타휴리스틱 알고리즘의 융합에 대한 논의
- 현재 연구 문제점/한계 분석 RL 기반 스케줄링 연구와 응용을 위한 연구 방향 제시

## 연구논문 vs 리뷰논문

본 논문은 리뷰 논문으로 생산 스케줄링에 대해 RL 기반의 최적화에 대한 연구 흐름과 분야별 주요 논문들을 요약한 논문입니다.

항 목	연구논문	리뷰논문
목적	어떤 주제에 대한 고유하고 구체적인 연구를 상세히 보고하는 것	특정 주제에 대해 출판된 문헌을 비평하고 분석하는 것
기본요소	항상 독창적인 연구 작업에 기초해야 하며, 주제에 대한 주요 참고 자료가 되어야 한다.	항상 발표된 학술적 문헌에 기초해야 하며 주제에 대한 새로운 정보는 포함하지 않아야 한다.
내용	연구 논문의 원본 데이터의 분석과 해석에 기초해야 한다.	원본 연구 논문의 간단하고 간결하게 요약된 내용을 포함하고 있으며, 주제에 대한 개요의 역할을 해야 한다.
보고내용	연구를 위해 수행된 모든 단계를 보고하고 추상적이고 잘 조직된 가설, 배경 연구, 방법론, 결론 및 연구 결과를 설명	주제에 대한 다양한 연구들 간의 공통점 뿐만아니라 상반되거나 다른 결과의 이유와의 불일치를 보고
길이	저널 출판이나 교육기관에 따라 다를 수 있지만 3000~6000단어까지 다양할 수 있다. 일부 저널은 더 긴 논문을 허용하기도 한다.	일반적으로3000~5000단어의 한도를 가지지만 논문의 장점에 따라 더 짧아질 수 있다.



분야나 논문마다 다르겠지만 리뷰 논문이다 일반적인 연구 논문에 비해 참고문헌의 수가 조금 많은 80여편에 이르고 있습니다. 특정 분야나 주제에 대한 논문을 리뷰해서 정리하는 논문이다 보니 그렇습니다.

comprehensively developed and enhanced.

## Acknowledgment

This work was supported in part by the National Science Fund for Distinguished Young Scholars of China (No. 61525304) and the National Natural Science Foundation of China (No. 61873328).

## References

- [1] N. Dilekhanakul, C. Kaplanis, N. Pawlowski, and M. Shanahan, Feature control as intrinsic motivation for hierarchical reinforcement learning, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3409–3418, 2019.
- [2] Y. Y. Jia and S. G. Ma, A coach-based bayesian reinforcement learning method for snake robot control, *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2319–2326, 2021.
- [3] I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, Neural combinatorial optimization with reinforcement learning, in *Proc. 3rd Int. Conf. Learning Representations*, Toulon, France, 2017, pp. 1–13.
- [4] W. Kool, H. Van Hoof, and M. Welling, Attention, learn to solve routing problems, in *Proc. 7th Int. Conf. Learning Representations*, New Orleans, LA, USA, 2019, pp. 1–12.
- [5] L. Wang and Z. X. Pan, Scheduling optimization for flow-shop based on deep reinforcement learning and iterative greedy method, (in Chinese), *Control and Decision*, vol. 36, no. 11, pp. 2609–2617, 2021.
- [6] L. B. Wang, X. Hu, Y. Wang, S. J. Xu, S. J. Ma, K. X. Yang, Z. J. Liu, and W. D. Wang, Dynamic job-shop scheduling in smart manufacturing using deep reinforcement learning, *Comput. Nerv.*, vol. 190, p. 107969, 2021.
- [7] S. H. Qu, J. Wang, S. Govil, and J. O. Leckie, Optimized adaptive scheduling of a manufacturing process system with multi-skill workforce and multiple machine types: An ontology-based, multi-agent reinforcement learning approach, *Procedia Cirp*, vol. 57, pp. 55–60, 2016.
- [8] S. Luo, L. X. Zhang, and Y. S. Fan, Dynamic multi-objective scheduling for flexible job shop by deep reinforcement learning, *Comput. Ind. Eng.*, vol. 159, p. 107489, 2021.
- [9] Y. C. Wang and J. M. Usher, Application of reinforcement learning for agent-based production scheduling, *Eng. Appl. Artif. Intell.*, vol. 18, no. 1, pp. 73–82, 2005.
- [10] H. F. Wang, Q. Yan, and S. Z. Zhang, Integrated scheduling and flexible maintenance in deteriorating multi-state single machine system using a reinforcement learning approach, *Adv. Eng. Inform.*, vol. 49, p. 101339, 2021.
- [11] Y. J. Zhao, Y. H. Wang, J. Zhang, and H. X. Yu, Application of improved Q learning algorithm in job shop scheduling problem, (in Chinese), *Journal of System Simulation*, <https://kns.cnki.net/kcms/detail/11.3092.V.20210423.1822.002.html>, 2021.

- [12] C. Zhang, W. Song, Z. G. Cao, J. Zhang, P. S. Tan, and C. Xu, Learning to dispatch for job shop scheduling via deep reinforcement learning, *arXiv preprint arXiv: 2010.12367*, 2020.
- [13] B. A. Han and J. J. Yang, Research on adaptive job shop scheduling problems based on dueling double DQN, *IEEE Access*, vol. 8, pp. 186474–186495, 2020.
- [14] L. Hu, Z. Y. Liu, W. F. Hu, Y. Y. Wang, J. R. Tan, and F. Wu, Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network, *J. Manuf. Syst.*, vol. 55, pp. 1–14, 2020.
- [15] D. Y. Zhang and C. M. Ye, Reinforcement learning algorithm for permutation flow shop scheduling to minimize makespan, (in Chinese), *Comput. Syst. Appl.*, vol. 28, no. 12, pp. 195–199, 2019.
- [16] M. A. L. Silva, S. R. de Souza, M. J. F. Souza, and A. L. C. Bazzan, A reinforcement learning-based multi-agent framework applied for solving routing and scheduling problems, *Expert Syst. Appl.*, vol. 131, pp. 148–171, 2019.
- [17] C. C. Lin, D. J. Deng, Y. L. Chih, and H. T. Chiu, Smart manufacturing scheduling with edge computing using multiclass deep Q network, *IEEE Trans. Ind. Inform.*, vol. 15, no. 7, pp. 4276–4284, 2019.
- [18] S. L. Yang, Z. G. Xu, and J. Y. Wang, Intelligent decision-making of scheduling for dynamic permutation flowshop via deep reinforcement learning, *Sensors*, vol. 21, no. 3, p. 1019, 2021.
- [19] S. Luo, Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning, *Appl. Soft Comput.*, vol. 91, p. 106208, 2020.
- [20] A. M. Kintsakis, F. E. Psomopoulos, and P. A. Mitkas, Reinforcement learning based scheduling in a workflow management system, *Eng. Appl. Artif. Intell.*, vol. 81, pp. 94–106, 2019.
- [21] Y. Y. Li, F. Fadda, D. Manerba, R. Tadei, and O. Tero, Reinforcement learning algorithms for online single-machine scheduling, in *Proc. 2020 Federated Conf. Computer Science and Information Systems*, Sofia, Bulgaria, 2020, pp. 277–283.
- [22] R. S. Willem and K. Setiawan, Reinforcement learning combined with radial basis function neural network to solve Job-Shop scheduling problem, in *Proc. 2011 IEEE Int. Summer Conference of Asia Pacific Business Innovation and Technology Management*, Dalian, China, 2011, pp. 29–32.
- [23] K. Ariv, H. Stern, and Y. Edan, Collaborative reinforcement learning for a two-robot job transfer flow-shop scheduling problem, *Int. J. Prod. Res.*, vol. 54, no. 4, pp. 1196–1209, 2016.
- [24] I. B. Park, J. Huh, J. Kim, and J. Park, A reinforcement learning approach to robust scheduling of semiconductor manufacturing facilities, *IEEE Trans. Automat. Sci. Eng.*, vol. 17, no. 3, pp. 1420–1431, 2020.
- [25] J. Wang, J. Hu, G. Y. Min, W. H. Zhan, Q. Ni, and N. Georgalas, Computation offloading in multi-access edge computing using a deep sequential model based on

- rescheduling using deep reinforcement learning, *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 231–236, 2019.
- [26] Z. H. Qin, N. Li, X. T. Liu, X. L. Liu, Q. Tong, and X. H. Liu, Overview of research on model-free reinforcement learning, (in Chinese), *Computer Science*, vol. 48, no. 3, pp. 180–187, 2021.
- [27] J. Palombini, J. C. Barco, and E. Martinez, Generating rescheduling knowledge using reinforcement learning in a cognitive architecture, *arXiv preprint arXiv: 1805.04752*, 2018.
- [28] R. H. Chen, B. Yang, S. Li, and S. L. Wang, A self-learning genetic algorithm based on reinforcement learning for flexible job-shop scheduling problem, *Comput. Ind. Eng.*, vol. 149, p. 108778, 2020.
- [29] A. I. Orban, F. Pop, and I. Raicu, New scheduling approach using reinforcement learning for heterogeneous distributed systems, *J. Parallel Distrib. Comput.*, vol. 117, pp. 292–302, 2018.
- [30] N. Aissani, A. Bekrar, D. Trentesaux, and B. Beldjiali, Dynamic scheduling for multi-site companies: A decisional approach based on reinforcement multi-agent learning, *J. Intell. Manuf.*, vol. 23, no. 6, pp. 2513–2529, 2012.
- [31] W. Bouazza, Y. Salliez, and B. Beldjiali, A distributed approach solving partially flexible job-shop scheduling problem with a Q-learning effect, *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 15890–15895, 2017.
- [32] Y. F. Wang, Adaptive job shop scheduling strategy based on weighted Q-learning algorithm, *J. Intell. Manuf.*, vol. 31, no. 2, pp. 417–432, 2020.
- [33] N. Stricker, A. Kuhnle, R. Sturm, and S. Fries, Reinforcement learning for adaptive order dispatching in the semiconductor industry, *CIRP Annals*, vol. 67, no. 1, pp. 511–514, 2018.
- [34] H. X. Wang and H. S. Yan, An interoperable adaptive scheduling strategy for knowledgeable manufacturing based on SMGWQ-learning, *J. Intell. Manuf.*, vol. 27, no. 5, pp. 1085–1095, 2016.
- [35] H. X. Wang, H. S. Yan, and Z. F. Wang, Adaptive assembly scheduling of aero-engine based on double-layer Q-learning in knowledge manufacturing, (in Chinese), *Computer Integrated Manufacturing Systems*, vol. 20, no. 12, pp. 3000–3010, 2014.
- [36] J. W. Liu, F. Gao, and X. L. Luo, Survey of deep reinforcement learning based on value function and policy gradient, (in Chinese), *Chinese Journal of Computers*, vol. 42, no. 6, pp. 1406–1438, 2019.
- [37] B. Waschneck, A. Reichstaller, L. Belzner, T. Allenmüller, T. Bauernhansl, A. Knapp, and A. Kyek, Optimization of global production scheduling with deep reinforcement learning, *Procedia CIRP*, vol. 72, pp. 1264–1269, 2018.
- [38] H. Hu, X. L. Jia, Q. X. He, S. F. Fu, and K. Liu, Deep reinforcement learning based AGVs real-time scheduling with mixed rule for flexible shop floor in industry 4.0, *Comput. Ind. Eng.*, vol. 149, p. 106749, 2020.
- [39] J. A. Palombini and E. C. Martinez, Closed-loop

- rescheduling using deep reinforcement learning, *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 231–236, 2019.
- [40] Z. H. Qin, N. Li, X. T. Liu, X. L. Liu, Q. Tong, and X. H. Liu, Overview of research on model-free reinforcement learning, (in Chinese), *Computer Science*, vol. 48, no. 3, pp. 180–187, 2021.
- [41] A. Kuhnle, N. Röhrig, and G. Lanza, Autonomous order dispatching in the semiconductor industry using reinforcement learning, *Procedia CIRP*, vol. 79, pp. 391–396, 2019.
- [42] C. L. Liu, C. C. Chang, and C. J. Tseng, Actor-critic deep reinforcement learning for solving job shop scheduling problems, *IEEE Access*, vol. 8, pp. 71752–71762, 2020.
- [43] C. D. Hubbs, C. Li, N. V. Sahasidis, I. E. Grossmann, and J. M. Wassick, A deep reinforcement learning approach for chemical production scheduling, *Comput. Chem. Eng.*, vol. 141, p. 106982, 2020.
- [44] X. Y. Chen and Y. D. Tian, Learning to perform local rewriting for combinatorial optimization, *arXiv preprint arXiv: 1810.00337*, 2019.
- [45] J. Wang, X. P. Li, and X. Y. Zhu, Intelligent dynamic control of stochastic economic lot scheduling by agent-based reinforcement learning, *Int. J. Prod. Res.*, vol. 50, no. 16, pp. 4381–4395, 2012.
- [46] S. F. Xie, T. Zhang, and Q. Rose, Online single machine scheduling based on simulation and reinforcement learning, in *Proc. of the Simulation in Produktion und Logistik, Wissenschaftliche Schriften*, Auerbach, Germany, 2019, pp. 59–68.
- [47] S. J. Wang, S. Sun, B. H. Zhou, and L. F. Xi, Q-learning based dynamic single machine scheduling, (in Chinese), *Journal of Shanghai Jiaotong University*, vol. 41, no. 8, pp. 1227–1232 & 1243, 2007.
- [48] H. B. Yang, W. C. Li, and B. Wang, Joint optimization of preventive maintenance and production scheduling for multi-state production systems based on reinforcement learning, *Reliab. Eng. Syst. Saf.*, vol. 214, p. 107713, 2021.
- [49] H. B. Yang, L. Shen, M. Cheng, and L. F. Tao, Integrated optimization of scheduling and maintenance in multi-state production systems with deterioration effects, (in Chinese), *Computer Integrated Manufacturing Systems*, vol. 24, no. 1, pp. 80–88, 2018.
- [50] Y. C. Wang and J. M. Usher, Learning policies for single machine job dispatching, *Robot. Comput. Integr. Manuf.*, vol. 20, no. 6, pp. 553–562, 2004.
- [51] Z. C. Zhang, L. Zheng, and M. X. Wang, Dynamic parallel machine scheduling with mean weighted tardiness objective by Q-learning, *Int. J. Adv. Manuf. Technol.*, vol. 34, no. 9, pp. 968–980, 2007.
- [52] L. F. Zhou, L. Zhang, and B. K. P. Horn, Deep reinforcement learning-based dynamic scheduling in smart manufacturing, *Procedia CIRP*, vol. 93, pp. 383–388, 2020.
- [53] Z. C. Zhang, L. Zheng, and X. H. Wang, Parallel machines scheduling with reinforcement learning, (in Chinese), *Computer Integrated Manufacturing Systems*, vol. 13, no. 1, pp. 110–116, 2007.

- agent reinforcement learning for online scheduling in smart factories, *Robot. Comput. Integr. Manuf.*, vol. 72, p. 102022, 2021.
- [68] B. Wang, F. G. Liu, and W. W. Lin, Energy-efficient VM scheduling based on deep reinforcement learning, *Future Generation Computer Systems*, vol. 125, pp. 616–628, 2021.
- [69] Y. He, L. X. Wang, Y. F. Li, and Y. L. Wang, A scheduling method for reducing energy consumption of machining job shops considering the flexible process plan, (in Chinese), *Journal of Mechanical Engineering*, vol. 52, no. 19, pp. 168–179, 2016.
- [70] J. Hong and V. V. Parbhia, Distributed reinforcement learning control for batch sequencing and sizing in Just-In-Time manufacturing systems, *Appl. Intell.*, vol. 20, no. 1, pp. 71–87, 2004.
- [71] T. Zhou, D. B. Tang, H. H. Zhu, and L. P. Wang, Reinforcement learning with composite rewards for production scheduling in a smart factory, *IEEE Access*, vol. 9, pp. 752–766, 2020.
- [72] J. L. Yuan, M. C. Chen, T. Jiang, and C. Li, Multi-objective reinforcement learning job scheduling method using AHP fixed weight in heterogeneous cloud environment, (in Chinese), *Control and Decision*, doi: 10.13195/kjzj.2020.0911.
- [73] W. H. Zhan, C. B. Luo, J. Wang, C. Wang, G. Y. Min, H. C. Duan, and Q. X. Zhu, Deep reinforcement learning-based offloading scheduling for vehicular edge computing, *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5449–5465, 2020.
- [74] Y. X. Yang, J. Hu, D. Porter, T. Marek, K. Hefflin, and H. X. Kong, Deep reinforcement learning-based irrigation scheduling, *Trans. ASABE*, vol. 63, no. 3, pp. 549–556, 2020.
- [75] A. Mortazavi, A. A. Khamseh, and P. Azimi, Designing of an intelligent self-adaptive model for supply chain ordering management system, *Eng. Appl. Artif. Intell.*, vol. 37, pp. 207–220, 2015.
- [76] C. M. Xing and F. A. Liu, An adaptive particle swarm optimization based on reinforcement learning, (in Chinese), *Control and Decision*, vol. 26, no. 1, pp. 54–58, 2011.
- [77] J. J. Wang and L. Wang, A cooperative memetic algorithm with learning-based agent for energy-aware distributed hybrid flow-shop scheduling, *IEEE Trans. Evol. Comput.*, doi: 10.1109/TEVC.2021.3106168.
- [78] Z. P. Li, X. M. Wei, X. S. Jiang, and Y. W. Pang, A kind of reinforcement learning to improve genetic algorithm for multiagent task scheduling, *Mathematical Problems in Engineering*, vol. 2021, p. 1796296, 2021.
- [79] M. Alcaistro, D. Ferone, P. Festa, S. Fugaro, and T. Pastore, A reinforcement learning iterated local search for makespan minimization in distributed manufacturing machine scheduling problems, *Comput. Operat. Res.*, vol. 131, p. 105272, 2021.

1. Introduction
2. State and action designs for scheduling
  - 2.1 Designs of state for scheduling
  - 2.2 Designs of action for scheduling
3. RL-Based Algorithm for scheduling
  - 3.1 Value-based RL for scheduling
  - 3.2 Policy-based RL for scheduling
4. RL Applications for scheduling
  - 4.1 RL for single machine scheduling
  - 4.2 RL for parallel machine scheduling
  - 4.3 RL for flow shop scheduling
  - 4.4 RL for job shop scheduling
  - 4.5 RL for other scheduling problems
5. Integration of RL and Meta-Heuristic for scheduling
6. Discussion and Conclusion
  - Problem domain / algorithm domain / application domain



- 생산 스케줄링은 제조 시스템에서 핵심 구성 요소 중 하나임
- 생산 효율을 향상시키기 위해 스케줄링 알고리즘은 산업공학, 자동화, 경영과학 등 학제간 분야에서 항상 중요한 연구 주제였음
- 통상적인 생산 스케줄링 알고리즘에는 다음의 3가지로 구분할 수 있음
  - accurate algorithms : 이론상 최적의 솔루션을 보장하지만 Np-hard 특성으로 인한 대규모 문제 해결이 어려움
  - heuristics : 전역 최적화 관점 없이 효율적인 스케줄링 방법을 구성하기 위해 몇가지 규칙을 채택하지만 문제의 특정한 특성에 대한 깊은 이해가 필요함
  - meta-heuristics : 허용 가능한 시간 내에 우수한 스케줄링을 얻을 수 있지만 동시에 대규모 문제의 경우 시간이 많이 걸리고, 온라인 음식 배달과 같은 실시간 시나리오에는 적용이 어려움
- AI 기술의 발전으로 RL은 게임, 로봇제어와 같은 순차적 의사 결정 문제에 성공적으로 적용되어 왔음
- 최근 몇 년 동안 RL은 Vehicle Routing Problem(VRP)/Traveling Salesman Problem(TSP)와 같은 여러 조합 최적화 문제를 해결하는데 성공적으로 적용되었음
- 생산 일정 문제가 RL의 환경으로 간주될 수 있다고 가정하면 에이전트는 다수의 offline training을 통해 환경과의 상호 작용 뿐만 아니라 합리적인 행동 및 상태 설계를 통해 정책 또는 규칙을 학습할 수 있음



- 이러한 새로운 아이디어는 스케줄링 문제, 특히 실시간 요구 사항이 높은 불확실하고 동적인 문제를 해결하기 위한 새로운 접근 방식을 제공
- Scopus에서 "reinforcement learning"과 "shop Scheduling"을 주제로 검색하여 다음과 같은 통계 데이터를 산출하였으며 이는 2015년 이후 RL 기반의 생산 스케줄링 최적화에 대한 논문이 빠르게 증가하고 있음을 의미하며 인공지능을 융합하여 RL을 기반으로 하는 스케줄링 최적화가 관련 분야에서 새로운 주제가 되고 있음을 의미함

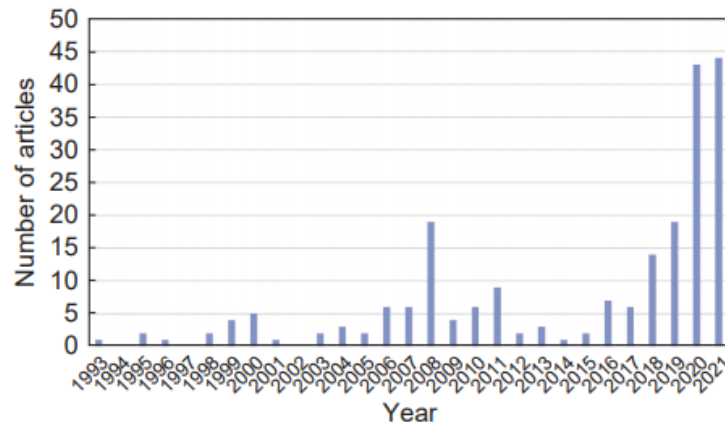


Fig. 1 Statistics of the RL-based scheduling in Scopus.

Table 1 Source of the articles.

No.	Publication	Number of articles
1	Lecture Notes in Computer Science	20
2	International J. of Production Research	9
3	Computers and Industrial Engineering	7
4	IEEE Access	7
5	IEEE Trans. on Automation Science and Engineering	5
6	Procedia CIRP	5
7	European Journal of Operational Research	4
8	Winter Simulation Conference	4
9	International J. of Advanced Manufacturing Technology	4
10	International J. of Simulation Modelling	3
11	IEEE Trans. on Industrial Informatics	3
12	Computer Integrated Manufacturing Systems	3
13	Advances in Intelligent Systems and Computing	3
14	Investigacion Operacional	3
15	Control and Decision	3
16	IEEE International Conference on Robotics and Automation	3

- RL의 중요한 구성 요소로서 합리적인 동작 및 상태 설계는 일정을 정확하게 설명하고 학습 프로세스의 효율성을 향상시킬 수 있음
- 각각 세가지 범주로 구분해 볼 수 있음
- Designs of state for scheduling
  - 생산 정보/생산 정보 통계
  - 정보의 양적 관계에 따른 방식
  - 그래프 활용
  - 기타
- Designs of action for scheduling
  - heuristic
  - job sequence
  - scheduling operators
  - 기타

- 가공정보, 가공 환경 정보, 주문 정보 등을 포함하는 생산 정보 또는 생산 정보 통계를 취하는 방식
- 정보 손실을 효과적으로 줄일 수 있으나 생산 정보는 일반적으로 연속적인 데이터 이기 때문에 문제 크기가 증가하면 차원 문제가 발생
- 따라서 신경망은 일반적으로 가치 함수와 정책 함수를 근사화 하는데 사용됨
- 관련 연구
  - the permutation flow shop scheduling(Wang and Pan)  
: 각 장비의 작업 처리 시간을 상태로 선택하고 정책을 학습하는 개선된 포인터 네트워크를 제안
  - the dynamic job-shop scheduling in smart manufacturing(Wang et al.)  
: 3가지(Job들에서 작업의 처리 시간, Job 처리 시간 그리고 Job 처리 상태) 매트릭스를 정의 하고 이를 입력값으로 사용하는 신경망을 통해 정책을 학습
  - 버퍼 크기, 워크스테이션 상태 정보, 인력 상태를 시스템의 상태로 설정하여 다음 액션 선택을 가이드 하는 방식(Qu et al.)
  - *the dynamic multi-objective flexible job shop scheduling problem(Luo et al.)*  
: *장비 대수, 장비 평균 가동률, 납품 일정 등 10가지 문제 정보를 기반으로 한 DQN(deep Q-learning Network) 제안*

- the dynamic multi-objective flexible job shop scheduling problem(Luo et al.)  
: 장비 대수, 장비 평균 가동률, 납품 일정 등 10가지 문제 정보를 기반으로 한  
DQN(deep Q-learning Network) 제안

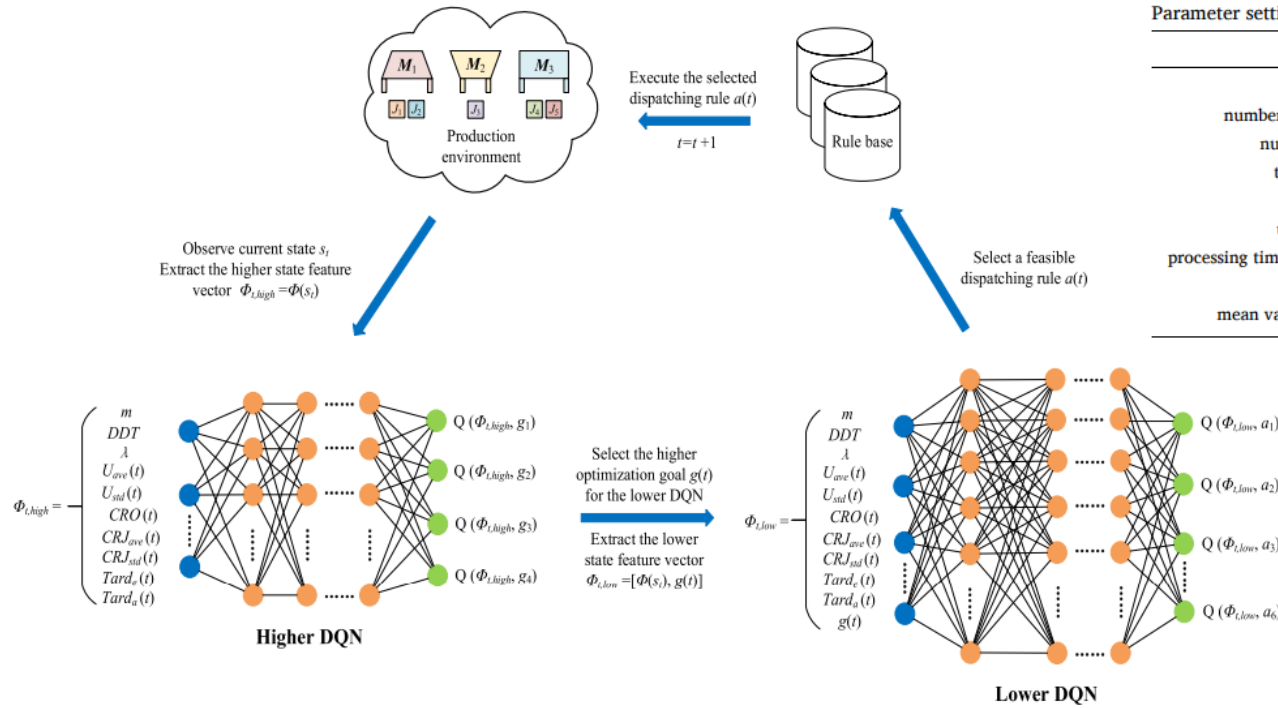


Table 2

Parameter settings for training and testing benchmarks.

Parameter	Value
total number of machines ( $m$ )	randi[10, 50]
number of available machines per operation	randi[1, $m$ ]
number of initial jobs at beginning	randi[1, 20]
total number new inserted jobs	randi[50, 200]
number of operations per job	randi[1, 20]
urgency degree $Pr_i$ of each job	randi[1, 5]
processing time of an operation on each available machine	randf[1, 50]
due date tightness ( $DDT$ )	randf[0.5, 1.5]
mean value $\lambda$ of the interarrival time $\exp(1/\lambda)$	randf[50, 200]

Fig. 1. Structure of the proposed THDQN.



- 생산 정보 또는 생산 정보 통계 간의 양적 관계에 따른 상태를 정의
- 이 문제 크기 증가로 인한 상태 공간 문제는 피할 수 있지만 정보 손실을 유발함
- 관련 연구
  - the dynamic single machine scheduling problem(Wang and Usher)
    - : 버퍼에 있는 작업 수에 대한 수량 상황과 총 지연 추정에 따라 상태를 정의하여 상태 공간을 효과적으로 줄임
  - the integrated scheduling and flexible maintenance in deteriorating multi-state single machine system(Wang et al.)
    - : 평균 정상 처리 시간과 남은 작업의 평균 지연 추정 간의 양적 관계 따라 상태 공간을 나눔
  - Application of improved Q learning algorithm in job shop scheduling problem (Zhao et al.)
    - : 예상 평균 여유 시간과 예상 평균 잔여 시간을 고려하여 6가지 상태를 정의

- 스케줄링 문제를 그래프로 변환하고 그래프의 노드와 에지의 상황에 따라 상태를 정의
- 이 방식은 문제의 구조적 특성을 잘 고려하여 생산 환경을 효율적으로 나타낼 수 있는 특징을 가짐
- 문제 특성을 효과적으로 추출하기 위해 일반적으로 GNN(Graph Neural Network), CNN(Convolutional Neural Network), GCN(Graph Convolutional Network) 및 기타 네트워크를 채택
- 관련 연구
  - job-shop scheduling 문제를 모델링하기 위해 분리형 그래프를 채택하고 GNN을 최적화하기 위해 PPO(Proximate Policy Optimization)를 제안(Zhang et al.)
  - *Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network(Hu et al.)*  
: 생산 정보를 다중 채널 이미지로 표현하고 CNN을 사용해서 state-action value 함수를 근사화 하고 문제를 모델링하기 위해 Petri-net을 사용했고 DQN에서 state-action value 함수를 근사화 하기 위해 GCN을 채택

- the permutation flow shop scheduling problem(Zhang and Ye)
  - 첫 번째 장비에 설정된 처리되지 않은 작업을 상태로 채택
  - 처리되지 않은 작업이  $n$ 개 있을 때 상태의 수는  $2^n$
- the unrelated parallel machine scheduling problem with sequence dependent setup time(Silva et al.)
  - 다중 에이전트 최적화 프레임워크를 제안하고 알고리즘의 상태로 4개의 이웃 구조를 설계
- 상태 디자인에는 여러가지 방법이 있을 수 있음
- 탁월한 상태 설계는 정보 손실과 상태 공간 크기의 균형을 맞춰야 하며 스케줄링 문제의 특성과 최적화 목표까지도 고려되어야 함

- 휴리스틱을 협력적으로 사용할 수 있음
- 액션의 수는 일정하고 문제의 크기와 독립적이나 알고리즘의 성능은 선택한 휴리스틱의 효율성과 품질에 따라 다름
- a smart manufacturing factory framework based on edge computing(Lin et al.)
  - Most-Operations-Remaining(MOR), First-In, First-Out(FIFO), Longest-Processing-Time(LPT)와 같은 7가지 휴리스틱을 DQN의 액션으로 선택
- the dynamic permutation flow shop scheduling problem(Yang et al.)
  - Shortest-Processing Time (SPT), LPT와 같은 5가지 규칙을 에이전트 액션으로 사용
- the dynamic flexible job shop scheduling with new job insertions(Luo)
  - 6개의 스케줄링 룰을 설계하고 규칙으로 사용



- a smart manufacturing factory framework based on edge computing(Lin et al.)
- Most-Operations-Remaining(MOR), First-In, First-Out(FIFO), Longest-Processing-Time(LPT)와 같은 7가지 휴리스틱을 DQN의 액션으로 선택

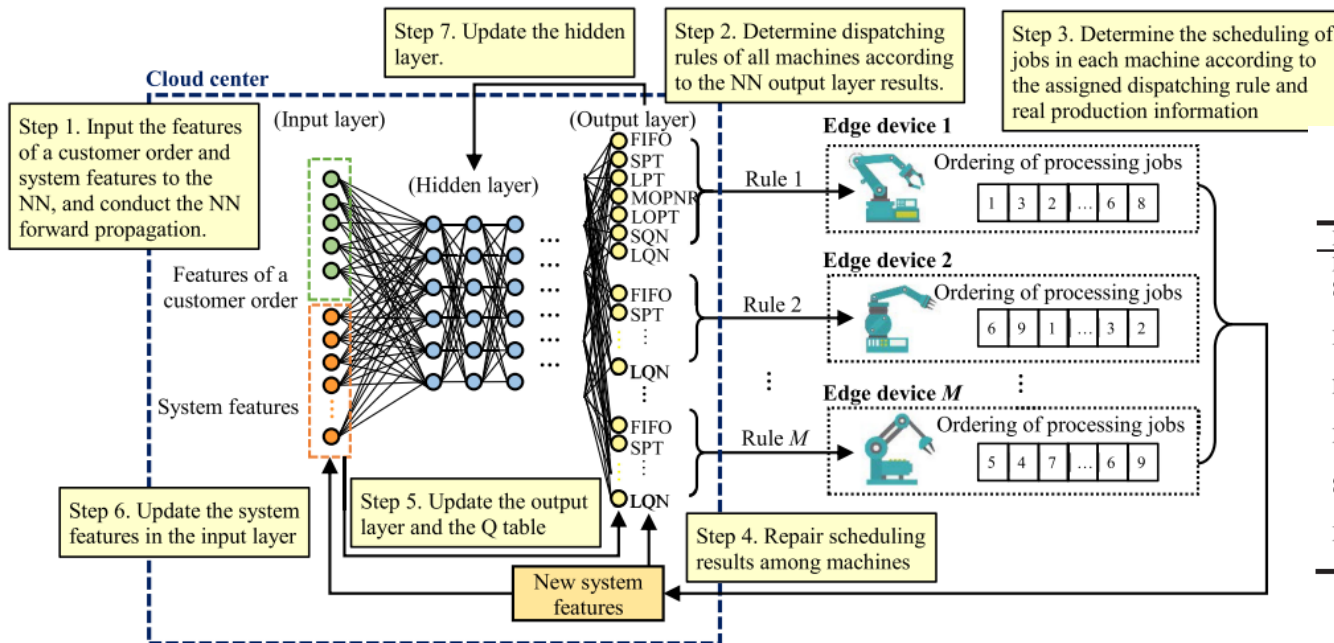


TABLE I  
DISPATCHING RULES CONSIDERED IN THIS PAPER

Dispatching rule	Content
First in first out (FIFO)	The first job is processed first.
Shortest processing time (SPT)	The job with the shortest processing time is process first.
Longest processing time (LPT)	The job with longest processing time is processing first.
Most operations remaining (MOPNR)	The job with most operations remaining at later tasks is processed first.
Longest operation processing time (LOPT)	The job with the longest processing time at later tasks is processed first.
Shortest next queue (SQN)	The job whose next task has the shortest processing time is process first.
Longest next queue (LQN)	The job whose next task has the longest processing time is process first.

Flowchart of the proposed MDQN method for the smart factory framework with edge computing.

- Job Sequence와 같은 것을 스케줄링 방식으로 사용
- 주로 End-to-End 모드를 사용하여 정적 스케줄링 문제를 해결하는 방식으로 에이전트는 스케줄링 방안을 빨리 수립할 수 있음
- the permutation flow shop scheduling(Wang and Pan)
  - 정책 네트워크를 통한 방식으로 프로세싱 정보를 사용하여 스케줄링 순서를 직접 출력함
- the workflow management system(Kintsakis et al.)
  - Sequence-to-Sequence 생성을 달성하기 위한 신경망을 디자인한 것으로 스케줄링 방안을 직접 출력

- 액션으로 문제들의 특징에 기반한 스케줄링 작업들을 정의
- 에이전트는 장비 할당 결정, 작업 순서 조정 등과 같은 적절한 작업을 선택하여 새로운 스케줄링 방안을 생성하는 방법을 학습함
- 이런 접근 방법은 실행 불가능한 방안을 생성하는 것을 방지하기 위해 문제의 특성 파악이 중요
- the online single machine scheduling(Li et al.)
  - 대기 큐에 있는 작업의 길이를 상태로 취하고 처리되지 않은 작업의 선택을 액션으로 정의함
  - Q-learning, SARSA(Single-Step State-Action-Reward-State-Action), 다단계 Watkins의 Q 및 다단계 SARSA를 각각 채택하여 문제를 해결
- the job shop scheduling problems(Williem and Setiawan)
  - 초기 상태로 critical path schedule을 선택, 풀 재할당과 작업 이동을 상태로 디자인 함
- the flow shop scheduling with two robot job transfer(Arviv et al.)
  - 2개의 Q-learning RL 알고리즘 제안
  - 작업 이동은 액션으로 정의되었으며 로봇과 생산 라인 간의 협력 스케줄링이 실현
- *the robust scheduling of semiconductor manufacturing facilities(Park et al.)*
  - 설비의 네 가지 로컬 기능을 연결하여 상태를 구성, 처리되지 않은 작업의 선택을 작업으로 정의

- *the robust scheduling of semiconductor manufacturing facilities(Park et al.)*
  - 설비의 네 가지 로컬 기능을 연결하여 상태를 구성, 처리되지 않은 작업의 선택을 작업으로 정의

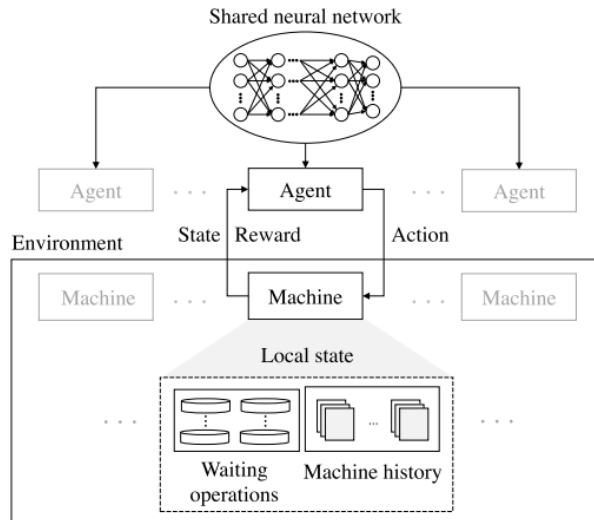


Fig. 2. Agent-oriented view of the proposed RL architecture.

TABLE II  
CONFIGURATION FOR THE EXAMPLE

Job types	Operations	Alternative machines	Initial setup status	$P(J_j)$
$J_1$	$O_{1,1}$	$M_1$	$O_{1,1}$	1
	$O_{1,2}$	$M_2$	-	
	$O_{1,3}$	$M_1, M_2$	$O_{1,2}$	
$J_2$	$O_{2,1}$	$M_1$	-	1

TABLE I  
COMPONENTS OF A STATE

Features	Descriptions	Dimension
Waiting operations	The number of waiting operations of $O_{j,k}$ which can be processed by the machine	$N_O$
Setup status	Setup type of the machine represented as one-hot encoding	$N_O$
Action history	The number of performed actions on the machine	$N_O + 1$
Utilization history	The amounts of processing, setup, and idle time of the machine	3

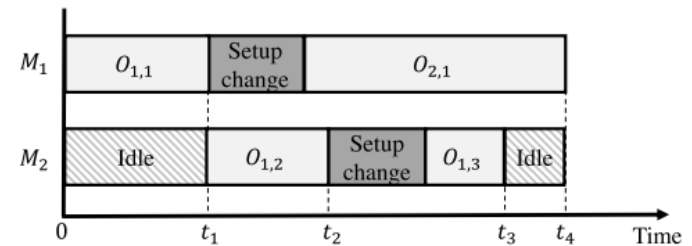


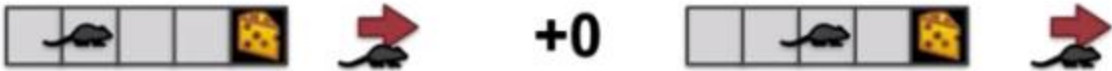
Fig. 3. Schedule obtained from the example.



- 위의 세 가지 범주 외에도 액션 디자인을 위한 다른 방법이 있음
- 적절한 형식과 액션 수를 생성하기 위해 문제의 속성을 고려해야 함
- a multi-agent framework combined with metaheuristics for unrelated parallel machines scheduling(Silva et al.)
  - 메타휴리스틱과 결합된 다중 에이전트 프레임워크를 제안, 액션은 이웃 구조의 선택으로 정의
- the scheduling problem in multi storage edge computing(Wang et al.)
  - PPO를 채택하고 작업을 실행할 위치를 선택하는 것을 작업으로 정의

- 환경 모델의 사용에 따라 RL은 모델 프리 RL과 모델 기반 RL의 두 가지 범주로 나눌 수 있음
- 모델 기반 RL은 상태 전환 및 보상 예측을 포함하는 환경 모델에 의존하며 에이전트는 새로운 상태와 보상을 직접 얻을 수 있지만 생산 일정 문제에 대한 상태 전이 정보를 얻기가 어려움
- 모델 프리 RL은 상태 전환 정보 없이 에이전트와 환경 간의 실시간 상호 작용에 기반하며 **현재 대부분의 기존 RL 기반 생산 일정 최적화 알고리즘은 모델 프리 RL 알고리즘으로 이를 다시 가치 기반 RL과 정책 기반 RL로 나눌 수 있음**
- 가치 기반(Value-based) RL
  - State-Action 값이 최대인 액션을 선택하여 최적의 전략을 구성하는 방식으로 가치 함수의 구성과 계산이 핵심이며 표본 활용도가 높지만 일반화 불량으로 과적합 되기 쉬움
  - 대표적 알고리즘으로 SARSA, Q-learning, DQN 등이 있음
- 정책 기반(Policy-based) RL
  - 가치 함수를 고려하지 않고 직접 최선의 정책을 찾는 방식으로 일반적으로 정책 기능에 맞게 신경망을 채택
  - 고유한 탐색 메커니즘이 있지만 샘플 활용률이 낮고 분산이 큰 로컬 최적화를 일으키기 쉬운 단점이 있으며 아직까지 정책 기반 RL은 스케줄링 문제에 널리 사용되지 않음
  - 대표적 알고리즘으로 REINFORCE, PPO 및 TRPO(Trust Region Policy Optimization) 등이 있음

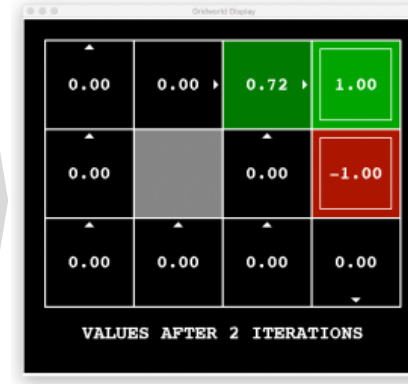
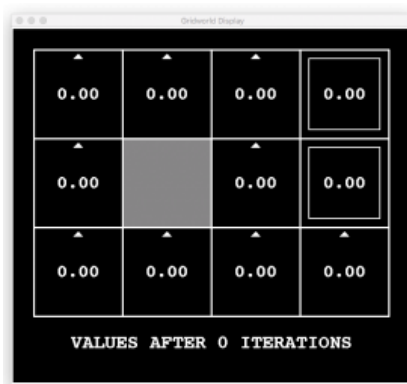
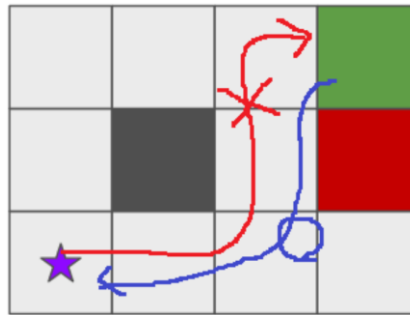
- SARSA 알고리즘의 이름은 강화학습의 각 원소에서 따온 것으로 시간차 학습을 사용해 상태와 행동 페어를 예측하는 방식

$$S_t \quad A_t \quad R_{t+1} \quad S_{t+1} \quad A_{t+1} \sim \pi$$

$$q(s_t, a_t) = q(s_t, a_t) + \alpha \times (r_{t+1} + \gamma \times q(s_{t+1}, a_{t+1}) - q(s_t, a_t)) \quad (1)$$

- 이를 적용한 다양한 연구 사례
  - 스케줄링 최적화를 위해 SARSA를 기반으로 일정 조정 지식을 생성하는 새로운 접근 방식을 제안하고 알고리즘의 효율성을 검증하기 위해 산업 사례를 테스트(Palombarini et al.)
  - 자가 학습 유전자 알고리즘(GA)을 설계하고 SARSA 및 Q-러닝을 도입하여 다양한 단계에서 검색 기능을 향상(Chen et al.)
  - 이기종 분산 시스템의 스케줄링 문제를 해결하기 위해 SARSA를 채택, 실험에 따르면 이 문제에서 SARSA의 성능이 Q-Learning보다 우수(Orhean et al.)

## Value-based RL for scheduling – Q-learning

- 다음 상태의 Q 값을 알고 있다는 전제하에 역산하여 내려오는 방식으로 시작 시점에는 최종 목적지의 보상값 밖에 알지 못하기에, 목적지부터 시작점까지 거꾸로 한 칸씩 Q값들을 계산하며 내려오는 원리



$$Q(s, a) = r(s, a) + \gamma \max_a Q(s', a)$$

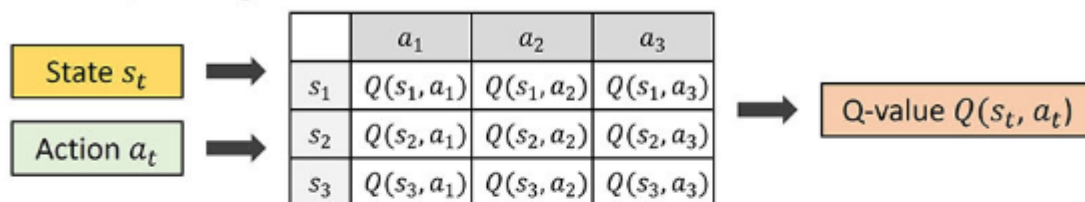


- Q-learning은 최근 몇년간 스케줄 최적화를 위해 진전된 모습을 보임
- flexible job-shop scheduling(Bouazza et al.)
  - Q-러닝을 적용하여 SQ(Shortest-Queue) 및 LQE(Less-Queued-Element)와 같은 규칙을 적용하여 장비 선택을 실현
  - FIFO 및 SJF(Shortest-Job-First)와 같은 일부 Rule이 작업 순서를 지정하기 위해 적용되었고 State-Action을 기록하기 위해 두 개의 Q-매트릭스를 사용함
- the dynamic job shop scheduling(Wang)
  - 클러스터링 및 동적 검색을 기반으로 하는 가중 Q-학습 알고리즘을 제안
  - 상태 공간의 차원을 줄이기 위해 네 가지 state function이 정의 되었고, 최적의 state-action 쌍을 선택하기 위해 개선된 반복 업데이트 전략을 제안
- *adaptive order dispatching in the semiconductor industry(Stricker et al.)*
  - *Q-learning을 사용하여 RL 기반 적응 제어 시스템을 설계*

## Value-based RL for scheduling – DQN(Deep Q-learning Network)

- DQN은 Google DeepMind에서 발표한 논문으로 딥러닝과 강화학습을 결합하여 인간 수준의 높은 성능을 달성한 첫번째 알고리즘
- Q-learning은 State-Action( $s,a$ )에 해당하는 Q-value인  $Q(s,a)$ 를 테이블 형식으로 저장하여 학습을 하는데 state space와 action space가 커지게 되면 시간/공간 복잡도가 증가하는 문제 발생
- DQN은 딥러닝을 이용하여 Q-table에 해당하는 Q-function을 비선형 함수로 근사해서 처리함

Classic Q-learning



Deep Q-learning



- *scheduling optimization(Waschneck et al.)*
  - DQN을 이용한 생산 스케줄링을 위한 DRL(Deep RL) 방법을 제시하고 제안된 알고리즘을 검증하기 위해 반도체 생산 사례를 채택
- *the flexible shop floor(Hu et al.)*
  - DQN을 사용하여 적응형 DRL 기반 AGV 실시간 스케줄링 접근 방식을 제안
  - 적절한 스케줄링 규칙과 AGV는 다양한 상태에서 선택할 수 있음
  - 실제 사례를 사용하여 알고리즘의 효율성을 검증

- 세 가지 가치 기반 RL 알고리즘 중 Q-learning은 욕심이 많고 로컬 최적화에 쉽게 갇히게 됨
- SARSA는 비교적 보수적이지만 수렴을 보장하기 위해  $\epsilon$ -greedy 방법의 탐사율을 제어해야 함
- DQN은 대규모 문제를 해결하는 데 적합하지만 DQN의 샘플링은 비효율적이며 매개변수 설정에 크게 의존함

- 정책기반 RL은 가치기반 RL과 달리 가치함수를 고려하지 않고 직접 최선의 정책을 찾는 방식으로 정책 기반에 맞게 신경망을 채택함
- 이러한 종류의 알고리즘에는 고유한 탐색 메커니즘이 있지만 샘플 활용이 낮고 분산이 큰 로컬 최적화를 일으키기 쉬움
- 다음 그림은 스케줄링을 위한 RL 기반 알고리즘의 통계를 나타내는 것으로 **생산 스케줄링을 해결하기 위해 가치 기반 RL에 대한 많은 연구가 있지만 정책 기반 RL에 대한 연구도 여전히 일정부분의 포션을 차지하고 있음**

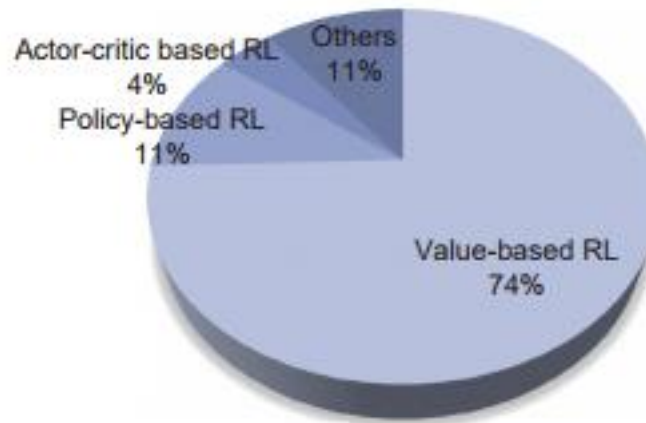


Fig. 3 Statistics of the RL-based algorithm for scheduling.

- permutation flow-shop scheduling problem(Wang and Pan)
  - 새로운 포인터 네트워크를 제안하고 네트워크를 훈련하기 위해 REINFORCE 방법을 채택하였고 알고리즘의 우수한 성능은 벤치마크에서 다른 알고리즘과 비교하여 입증됨
- Rummukainen과 Nurminen[40]은 PPO를 적용하여 확률적 경제적 Lot 스케줄링 문제를 해결
- Zhang et al.[12] Job-shop 스케줄링을 설명하고 GNN에 입력하기 위해 분리형 그래프를 사용했음 PPO는 네트워크를 훈련하는 데 사용되고 실험을 통해 알고리즘의 성능이 우수한 것으로 나타났음
- 더 작은 배치 크기, 더 큰 제품 다양성 및 생산 시스템의 복잡한 프로세스의 특성을 고려하여 TRPO를 기반으로 한 자율 디스패치 알고리즘을 설계했고, 제안된 스케줄링 알고리즘의 효율성은 반도체 산업의 실제 사례를 이용하여 검증하였음(Kuhnle et al.[41])
- 세 가지 정책 기반 RL 방법 중 REINFORCE는 Policy gradient 기반 Monte Carlo 알고리즘에 속하며 안정성은 좋지만 샘플 활용도가 낮음
- PPO 및 TRPO의 성능은 수퍼 매개변수에 크게 의존하지 않지만 강력한 실행 환경 지원으로 샘플링 속도가 낮음

- 가치 기반/정책 기반 RL 외에도 actor-critic 방식과 같은 다른 유형의 RL 방법이 있음
- Job-shop 스케줄링 문제를 해결하기 위해 비동기식 업데이트와 DDPG를 사용하여 모델을 학습시키는 병렬 학습 방법을 제안(Liu et al.[42])
- 화학 생산 일정을 해결하기 위해 advantage actor-critic 방식을 적용, 실험은 RL의 속도와 유연성이 스케줄링 시스템의 실시간 최적화를 실현하는 데 도움이 된다는 것을 증명함(Hubbs et al.[43])
- 온라인 작업 스케줄링을 위한 정책을 학습하기 위해 NeuRewriter를 제안하고 신경망을 훈련하기 위해 Actor-Critic 방법을 사용 했으며 실험 결과 제안한 알고리즘의 효율성을 확인함(Chen and Tian[44])



- 아래 그림은 다양한 유형의 스케줄링 문제에 대한 RL의 적용에 대한 논문 수를 보여줌
- RL은 주로 Job-shop 스케줄링 문제를 해결하는 데 사용됨을 알 수 있음
- Flow-shop, 병렬 기계 및 단일 기계 스케줄링 문제에 대한 application은 추가 연구가 필요함(연구가 아직 미진한 부분 또는 RL 기반의 연구 필요성이 낮거나)

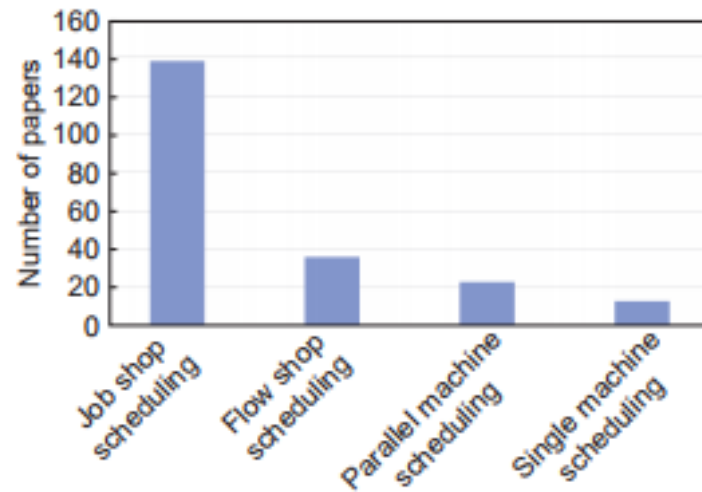


Fig. 4 Number of papers of RL in different types of scheduling.

- 단일 기계 스케줄링의 제약은 비교적 간단하며 작업의 작업 프로세스 순서만 결정하면 되는 것으로 현재 RL은 확률적, 동적 또는 온라인 조건에서 단일 기계 스케줄링 문제를 해결하는 데 주로 사용됨
- 단일 기계의 재고 생산 시스템에 대한 확률론적 경제적 Lot 스케줄링 문제를 해결하기 위해 두 가지 RL 기반 방법을 설계하고 비교했음(Wang et al.[45])
- 다중 상태 생산 시스템의 생산 일정 및 예방 유지 관리를 문제를 Markov 결정 프로세스로 변환하고 문제를 해결하기 위해 모델이 없는 RL 알고리즘을 설계함(Yang et al.[49])
- 작업 도착 시간이 Poisson 분포를 따른다는 점을 고려하여 단일 기계 스케줄링 문제에 대한 평균 지연 시간을 최소화하기 위해 세 가지 스케줄링 규칙을 동적으로 선택하는 Q-러닝을 제안(Wang and Usher[50])

- 단일 기계 스케줄링과 비교하여 병렬 기계 스케줄링은 기계 할당을 고려해야 하며 에이전트 상태와 동작이 복잡
- RL 기반 병렬 기계 스케줄링 최적화 알고리즘은 주로 동적 스케줄링 문제를 위해 설계됨
- 시퀀스 종속 설정 시간 및 기계 작업 자격 고려 사항이 있는 동적 병렬 기계 스케줄링 문제의 경우 평균 가중 지각을 최소화하기 위해 Q-러닝을 채택하고 5가지 휴리스틱을 행동으로 선택(Zhang et al.[51])
- Smart Manufacturing 스케줄링 문제에 대해 최대 완료 시간을 최소화하기 위해 Deep RL 기반 방법을 제안하였고 목표 네트워크와 예측 네트워크는 학습 과정에서 협력하여 안정성을 향상시키는 데 사용됨(Zhou et al.[52])

- Flow Shop 스케줄링은 여러 단계의 처리를 고려해야 함
- 유연한 제조를 실현하기 위해 하이브리드 또는 유연한 flow Shop 스케줄링과 같은 일부 단계에서 여러 병렬 기계가 있음
- 병렬 기계 스케줄링보다 더 복잡한 구조를 가짐
- 순열 flow Shop 스케줄링을 순차 결정 문제로 변환하고 Q-러닝 기반 스케줄링 알고리즘을 제안하여. 벤치마크 인스턴스를 사용하여 알고리즘의 효율성을 검증했음(Zhang and Ye[15])
- 시퀀스 종속 설정 시간이 있는 flow Shop 스케줄링에 대해 모든 작업의 완료 시간을 최소화하기 위해 개선된 Q-러닝을 제시(Fonseca-Reyna and Martínez-Jiménez[58])

- 앞선 세 가지 스케줄링 문제와 비교하여 Job-shop 스케줄링은 작업에 대한 다른 기계 처리 경로를 고려해야 하는데 유연한 스케줄링을 위해서는 기계 할당도 고려해야 하기 때문에 스케줄링 알고리즘의 설계는 더 복잡함
- 정적 시나리오에서 전통적인 Job-shop 스케줄링을 순차적 결정 문제로 변형하고 가치 함수를 근사화하기 위해 신경망을 도입하였고 시뮬레이션 결과, 설계된 알고리즘의 성능이 기존 규칙보다 우수한 것으로 나타났음(Gabel and Riedmiller[60])
- 새로운 작업 삽입과 관련된 유연한 작업장 일정 문제에 대해 6가지 파견 규칙을 제안하고 전체 지연을 최소화하기 위해 심층 Q-네트워크를 개발(Luo[19])
- 무작위 작업 도착 및 기계 고장과 관련된 동적 작업장 스케줄링 문제의 경우 평균 흐름 시간을 최소화하기 위해 RL 기반 변수 이웃 탐색을 제안했음, 작업 수와 현재 작업의 평균 처리 시간을 사용하여 여러 상태를 정의하였으며 Q-learning은 다른 상태에서 매개변수 선택을 학습하는 데 사용하였음(Shahrabi et al.[64])

- RL은 분산 스케줄링, 에너지 효율 스케줄링 및 다중 목표 스케줄링과 같은 일부 다른 유형의 스케줄링 문제에도 적용되었고 에지 컴퓨팅 작업 일정 및 농업 관개 일정과 같은 몇 가지 실제 생산 시나리오에서 진전을 보이기도 했음
- 분산 시스템에서 고차원 데이터를 다루기 온라인 스케줄링을 위한 스마트 공장의 새로운 사이버-물리적 통합 방법을 제시하였는데 RL은 스케줄링 알고리즘의 의사결정 능력을 향상시키기 위해 도입하였음(Zhou et al.[67])
- 가상 머신의 에너지 효율성 스케줄링 문제에 대해 서비스 품질(QoS) 기능 학습을 기반으로 하는 심층 RL 모델을 제안했고, 광범위한 실험을 통해 제안된 방법이 에너지 소비를 효과적으로 줄일 수 있음을 보여줌(Wang et al.[68])
- *Just-in-time 제약이 있는 동적 다중 목표 Job Shop 스케줄링 문제에 대해 문제를 SMDP로 모델링하고 Q-러닝을 사용하여 새로운 스케줄링 알고리즘을 도입하였고 알고리즘의 성능은 다른 스케줄링 규칙보다 훨씬 우수했음(Hong and Prabhu[70])*
- *반도체 산업의 다목적 스케줄링을 위해 가중 방법을 사용하여 두 가지 목표 최적화 문제를 처리했는데 RL을 도입하면 스케줄링 솔루션이 자동으로 생성될 수 있음(Kuhnle et al.[41])*
- 이기종 클라우드 환경에서 다중 목표 스케줄링 문제에 대해 실행 시간, 에너지 소비 및 실행 비용을 최적화하기 위해 분석 계층 프로세스를 기반으로 하는 다목적 강화 학습을 설계함(Yuan et al.[72])

- 정적 스케줄링과 동적 스케줄링에 대한 RL의 통계를 보면 **동적 스케줄링 문제를 해결하기 위해 RL 기반 생산 스케줄링 최적화가 주로 연구되었음을** 알 수 있음
- 그 이유는 환경과의 상호 작용에서 문제의 속성과 지식을 학습한 후 RL 에이전트가 동일한 문제 시나리오에서 다음 동적 스케줄링 문제를 계속 해결하여 솔루션을 빠르게 얻을 수 있기 때문으로 보임
- 그러나 정적 스케줄링 문제의 경우 훈련된 RL 에이전트가 다른 시나리오의 인스턴스를 해결하도록 확장될 때 RL로 얻은 솔루션은 종종 메타 휴리스틱으로 얻은 솔루션만큼 좋지 않는 경우 발생
- 따라서 정적 스케줄링을 위한 종단 간 모델을 사용한 RL에 대한 연구는 추가 연구가 필요함
- 또한, RL은 단일 목표라도 간단한 시나리오의 스케줄링 문제에 주로 적용되므로 복잡한 스케줄링 문제를 해결하기 위한 RL과 다중 목표 최적화에 대한 연구가 강조되어야 함

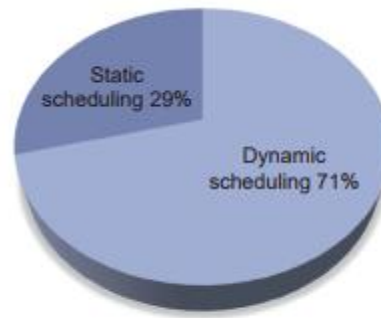


Fig. 5 Statistics of the RL for static scheduling and dynamic scheduling.



- 스케줄링을 위한 RL 애플리케이션은 매우 유망하며 여전히 논의되고 연구되어야 하는 분야임
- 최근 수십 년 동안 인공지능의 중요한 분야로서 계산 지능, 특히 메타 휴리스틱은 생산 일정에서 큰 발전을 이루었으나 단일 검색 모드의 메타 휴리스틱은 분산 스케줄링 및 그린 스케줄링과 같은 복잡한 스케줄링 문제를 효과적이고 효율적으로 처리하기 어려움
- 검색 효율성을 향상시키기 위해 메타 휴리스틱을 지원하기 위해 학습 메커니즘과 같은 다양한 메커니즘을 도입하는 것이 필요함
- 따라서 **RL과 메타 휴리스틱의 통합은 알고리즘 성능을 향상시키는 유망한 방법으로 볼 수 있음**

- 생산 일정은 제조 시스템의 핵심이며 많은 관심을 불러일으키는 분야임
- 대규모 및 실시간 요구 사항을 고려할 때 기존 스케줄링 알고리즘은 큰 도전에 직면해 있음
- 인공지능 기술의 발전으로 RL은 많은 조합 최적화 문제에서 획기적인 발전을 이루었으며 스케줄링 최적화를 위한 새로운 방법을 제공하고 있음
- 본 논문에서는 생산 스케줄링을 위한 RL 지능적 최적화를 위한 가이드라인을 제공하기 위해 생산 스케줄링을 위한 RL을 검토했음
- RL 기반 스케줄링에 대한 기존 연구에서 RL 알고리즘은 shop 스케줄링 문제, 특히 동적 스케줄링 문제를 해결하는 데 편리함과 신속성과 같은 특별한 이점이 있음을 확인할 수 있었음
- 그러나 아직까지도 관련 연구는 아직 초기 단계이며 문제, 알고리즘 및 응용 영역에서 더 많은 연구가 필요함

- 기존 작업은 주로 단일 목표 스케줄링 문제를 해결하기 위해 RL에 중점을 두고 있음
- 다목적 최적화에 대한 몇몇 연구는 주로 경제 및 시간 지수를 고려함 한편으로는 다양한 요구 사항에 따라 기계 부하 균형, 지연된 작업 수 및 기타 일정 지표를 연구해야 함
- 다른 한편으로, 탄소 피크 및 탄소 중립 목표의 제안은 산업의 녹색 변혁을 촉진하고 지능형 제조와 녹색 제조의 통합을 가속화하고 있음
- 따라서 경제적 목표와 녹색 목표를 동시에 최적화하기 위해 RL 알고리즘을 탐색하는 것은 실질적인 의미가 있음
- 또한 생산 일정 문제에 대한 RL에 대한 대부분의 문헌은 단순화되고 전통적임
- 동시에 유허 없음, 대기 없음, 시퀀스 종속 설정 시간 및 기계 성능 저하 효과와 같은 많은 실제 제약 조건을 고려해야 하며 복잡한 공정 제약이 있는 생산 일정 문제를 해결하기 위해 RL 알고리즘을 연구하는 것은 매우 실용적인 가치가 있음

- 현재 기존 RL 알고리즘은 스케줄링 문제를 해결하기 위한 이론적 분석 및 지원이 부족함
- 게다가, 상태와 행동의 설계를 안내하는 체계적인 방법의 부재도 생산 일정 문제를 해결하는 RL의 촉진 및 적용에 불리함
- 따라서 생산 일정 최적화를 위한 RL 알고리즘의 이론과 방법에 대한 연구는 매우 중요한 학문적 가치를 지닌다고 볼 수 있음
- 현재 생산 일정 문제에 거의 사용되지 않는 정책 기반 RL 알고리즘은 최적의 정책을 검색하고 종단간 방식으로 일정을 생성할 수 있으므로 실시간 시나리오의 문제에 효과적으로 대처할 수 있음
- 따라서 PPO 및 TRPO와 같은 정책 기반 RL 알고리즘의 연구를 통해 생산 일정 문제를 end-to-end 방식으로 해결하고 일정 규칙의 적응 생성을 실현하는 것이 중요한 것으로 보임
- 메타 휴리스틱과의 시너지를 고려할 때 협동적 RL에 대한 연구는 상대적으로 드문 분야임
- RL과 메타 휴리스틱의 효과적인 융합 메커니즘을 탐구하는 것은 유망한 연구 방향임
- 검색 방향 및 검색 단계 길이를 결정하고 검색 작업 및 매개변수 설정을 적응적으로 조정하는 RL의 장점을 최대한 활용하여 관련 지식을 찾고 검색 효율성을 향상시킬 것으로 예상됨

- 현재 스케줄링 문제에 대한 RL에 대한 대부분의 연구는 학문적 수준에 머물고 있음
- 관련 이론 및 방법은 실제 문제의 적용이 부족하여 시뮬레이션을 통해서만 테스트 및 분석되는 실정으로 실제 문제에 대한 이해와 정교화를 강화하고 문제 모델링 및 알고리즘 설계를 강조하며 Job-shop 스케줄링을 해결하기 위한 RL 알고리즘의 적용을 촉진할 필요가 있음
- 요컨대, RL을 기반으로 한 생산 일정 최적화에 대한 연구는 유망하지만 많은 영역이 개선되고 탐구되어야 하며 RL 기술의 발전으로 이론, 방법 및 응용 연구가 종합적으로 개발되고 향상될 수 있다고 예상됨



**감사합니다.**