

# Reinforcement Learning for Dynamic Job-Shop Scheduling with Setup Time Constraints\*

Kamil MUSIAŁ <sup>[0000-0002-7615-7165]</sup>, Artem BALASHOV <sup>[0000-0002-2937-1070]</sup>,  
Joanna KOCHAŃSKA <sup>[0000-0002-4695-1640]</sup> and Dagmara ŁAPCZYŃSKA <sup>[0000-0002-4745-0768]</sup>

Wroclaw University of Science and Technology, Wroclaw, Poland

Correspondence should be addressed to: Kamil MUSIAŁ, [kamil.musial@pwr.edu.pl](mailto:kamil.musial@pwr.edu.pl)

\* Presented at the 46<sup>th</sup> IBIMA International Conference, 26-27 November 2025, Ronda, Spain

## Abstract

Optimization in job-shop production systems is a very complex task, requiring simultaneous analysis of many factors. This article presents just this type of production in which the problem was to schedule job-shop tasks with significant setup times. In this case, the use of reinforcement learning (RL) was proposed, with particular emphasis on algorithm evaluation: Q-learning, actor-critic algorithms, deep neural networks (DQN), as well as advanced neural graph-based architectures (GNN). As a result of the analysis, it was shown that methods that are based on deep reinforcement learning (DRL) allow to achieve better results than traditional distracting heuristics and metaheuristic algorithms. Neural networks, which are based on heterogeneous graphs, are particularly useful for modeling relationships between machines and operations in the face of dynamically occurring changes in the production schedule. In this article, the methods described are analyzed in the context of production in small and medium-sized enterprises (SMEs), for which intensive technological development is a challenge. Ways of implementing RL solutions for such enterprises were also proposed.

**Keywords:** job-shop, optimization of production schedules, neural networks, Industry 4.0

## Introduction

Job-shop systems are characterized by unit production or in small batches, which can be carried out in a very diverse way, which requires a dynamic approach to introducing frequent and sudden changes in production. Therefore, planning and scheduling this type of production is a very complex task, especially in a rapidly changing environment and with high competition. Customers require personalization of orders and flexibility in making changes, which makes it difficult to plan the production process, especially with the use of traditional methods. This is even more difficult for processes that have long setup times or preparation and teardown times, which account for a significant portion of the total production time.

In order to remain competitive, companies must adapt to dynamically occurring changes and react flexibly to them by introducing changes in production schedules. The use of reinforcement learning allows for autonomous learning of optimal planning, while maintaining interaction with the production environment. In recent years, there has been an increase in interest in the use of DRLs for production scheduling and planning (Esto et al., 2022)

This article provides an overview of current solutions and the state of knowledge in the context of process planning in job-shop manufacturing. The performance of different RL algorithms in the context of application in SMEs was compared.

---

**Cite this Article as:** Kamil MUSIAŁ, Artem BALASHOV, Joanna KOCHAŃSKA and Dagmara ŁAPCZYŃSKA, Vol. 2025 (34) "Reinforcement Learning for Dynamic Job-Shop Scheduling with Setup Time Constraints " Communications of International Proceedings, Vol. 2025 (34), Article ID 4642925, <https://doi.org/10.5171/2025.4642925>

## **Job Shop Scheduling Problem (JSSP)**

The scheduling problem in the Job Shop Scheduling Problem (JSSP) system is to assign operations to machines and equipment in such a way as to achieve the assumed goal (e.g. minimize production time, maximize the use of resources, etc.). An additional aspect complicating this task is the occurrence of changeovers (i.e. activities necessary to change the type of manufactured element at a given position) and the dynamics of production processes, characteristic of job-shop systems – new tasks may appear during the process, and, as in any other process, unpredictable interruptions may occur, related to, for example, machines or changes in production priorities. Often, traditional methods are ineffective in such cases. All enterprises, including small and medium-sized ones, are forced to react quickly to changes in order to remain competitive on the market. With the development of technology, this is often associated with the need to make changes in the direction of meeting the assumptions of Industry 4.0.

## **Reinforcement Learning Algorithms for Job-Shop Planning**

Among the reinforcement learning methods that can be used to plan job-shop systems, several algorithms can be distinguished, which are:

### ***Q-Learning and Value-Based Algorithms***

Q-learning is a machine learning method in which an algorithm learns what decisions to make in different situations to achieve the best possible result. In job shop planning, Q-learning can be used to learn the so-called dispatch rules – i.e. the rules that tell which task to assign to which machine at a given time. If Q-learning properly presents the state of the system (e.g. by determining how "urgent" the tasks are not yet completed) and adds a penalty for too much time to the reward function, the algorithm can plan better than traditional, manually set rules.

The advantage of Q-learning is that it is easy to implement and does not require very much computing power. Its disadvantage is that it copes worse with big problems – when the number of possible states and decisions is very large, it is necessary to remember too much information. Therefore, in practice, more advanced versions of this algorithm are often sought (Coto et al., 2022).

### ***Deep Q-Networks (DQN)***

Deep Q-networks (DQNs) are an extension of the Q-learning method, in which instead of writing all possible results in a table, a neural network is used to approximate them. Thanks to this, the algorithm can work even when the number of possible situations (states) is very large.

In production planning, DQN proves to be very effective. For example, in the semiconductor industry, research has shown that DQN can create better task schedules than traditional scheduling rules. This is because the model can learn on its own what information about the state of the system is most important and how to react in different situations. The biggest advantage of DQN is that it learns autonomously from data, without the need for human input features. (Waschnech et al., 2018; Lee et al., 2021)

### ***Actor-Critic algorithms***

Actor-Critic algorithms are reinforcement learning methods that combine two strategies: one for making decisions (the so-called actor) and the other for evaluating those decisions (the so-called critic). The actor learns what actions to perform, and the critic evaluates how good those actions are and helps the actor improve.

In the planning of job-shop production, such algorithms have proven to be very effective. Specially designed versions for the so-called Flexible Job Shop Scheduling Problems (FJSSP) achieve better results than traditional planning rules or other reinforcement learning methods. Their big advantage is that they quickly adapt to changes in production conditions.

Some versions of Actor-Critic also use an attention mechanism that allows the model to focus on the most important elements of the situation — allowing them to make more accurate decisions in more complex environments (Monaci et al., 2024)

### ***Graph Neural Networks (GNN)***

Modern approaches to job-shop planning are increasingly using graph-based neural networks (GNNs). In this approach, the whole problem is presented as a heterogeneous graph, in which the vertices correspond to operations and machines and the edges describe the relationships between them.

GNNs are well suited for this type of task because they can model complex relationships between system components. Heterogeneous graph neural networks are particularly useful, as they can capture different types of connections – both between operations within a single task and between operations and machines. Research shows that methods based on GNN and reinforcement learning achieve better results in solving complex problems than traditional planning techniques.

A big advantage of GNNs is their ability to generalize – models trained on smaller examples often also cope well with larger or slightly different problems (Fernandez et al., 2025)

### ***Sequential Encoding-Decoding (Seq2Seq) Methods***

Another approach is models inspired by natural language processing solutions. In such methods, the list of operations is first encoded by the encoder, and then the decoder generates a step-by-step (token-by-token) task allocation policy. Such models are good at learning the general principles of dispatching, independent of the specific problem. Comparisons with classical heuristics show that the Seq2Seq approach often outperforms traditional priority rules used in task scheduling (Bonetta et al., 2023)

## **Comparison of RL Algorithm Performance**

The performance of planning algorithms is most often evaluated by:

- **Makespan (Cmaks)** - the total time of task completion (here: production),
- **Total Weighted Tardiness** - the sum of delays of tasks weighted by their weights (meaning),
- **Machine Utilization** - the degree of utilization of machines and equipment,
- **Average Lateness** - the average latency in task completion.

## **Comparative results**

On the basis of a review of the literature and our own analyses, it can be concluded that:

- **Q-Learning:** Suitable for simpler instances, requires significantly less computational time than more advanced methods.
- **DQN:** achieves a makespan reduction compared to traditional dispatching heuristics (such as the FIFO or SPT principle), but the computational requirement is much higher.
- **Actor-Critic:** Exhibits performance comparable to or better than DQN, while also having better training stability.
- **GNN + DRL:** achieves the best results, especially for large instances; makespan reduction to 20-35% and also maintains good generality on new instances.
- **Seq2Seq:** shows competitive results with GNN + DRL, with sometimes better generality on completely new instances of problems.

However, it should be remembered that more advanced and complicated methods (such as GNN or Seq2Seq) require much more data for training and computational time.

## **Practical Challenges and Limitations**

- **Generalization problem** – One of the main problems in teaching planning models is their limited ability to generalize. A model trained on data of a specific size (e.g., 10 jobs and 5 machines) often performs

worse when it encounters a problem of a different scale. GNN networks and Seq2Seq approaches have an advantage in this respect – they tend to transfer the acquired knowledge to new, larger cases better than classic DQN models.

- **Computational and Time Requirements** - training advanced reinforcement learning (RL) models, especially those based on GNN or Seq2Seq, is very expensive. It requires powerful graphics cards, a lot of RAM and often many days or even weeks of training. For smaller manufacturing companies (SMEs) that do not have such resources, simpler algorithms, such as classic Q-learning, can be much more practical.
- **Safety and Reliability in Real Systems** - Implementing RL models directly in real production facilities is risky. Even minor scheduling errors can lead to large losses or downtime. Therefore, safety mechanisms are needed – e.g. human supervision, the ability to return to proven heuristics and thorough testing in a simulation environment before implementation in real production.

## Summary

Reinforcement Learning (RL) is an effective tool for solving the problems of dynamic job-shop production planning, especially where changeover times between operations need to be taken into account. A review of the literature shows significant developments in this field, ranging from simple algorithms such as Q-learning to advanced architectures based on graphs (GNN) and sequential models (Seq2Seq).

The most important conclusions from the analysis:

1. Advanced architectures (GNN, Seq2Seq) achieve much better results than classic heuristics.
2. There is a trade-off between performance and computational complexity – the choice of method should depend on the capabilities and needs of a particular company.
3. The ability of models to generalize on new cases is crucial for practical applications.
4. Hybrid approaches, combining RL methods with traditional planning principles, offer the most realistic deployment path.

For small and medium-sized manufacturing companies that are developing towards Industry 4.0, reinforcement learning can become an important part of digital transformation. It allows to automate and optimize production planning, increasing efficiency and flexibility. In the future, research should focus on improving the generalization and safety of models, as these are the issues that still limit their wider use in industrial practice.

## References

- Bonetta, G., Zago, D., Cancelliere, R., Grosso, A. (2023). “Job Shop Scheduling via Deep Reinforcement Learning: a Sequence to Sequence approach”, Proceedings of the International Conference on Learning and Intelligent Optimization (LION17). Springer International Publishing.
- Coto Palacio, J., Martinez, Y., Schietgat, L., Doninck, B., Nowe, A. (2022). “A Q-Learning algorithm for flexible job shop scheduling in a real-world manufacturing scenario”, *Procedia CIRP*, 106. 227-232.
- Estes, A., Peidro, D., Mula, J., Díaz-Madroño, M. (2022). “Reinforcement learning applied to production planning and control”, *International Journal of Production Research*, 61. 1-18.
- Lee, Y., Lee, S. (2021). “Deep reinforcement learning based scheduling within production plan in semiconductor fabrication”, *Expert Systems with Applications*. 191. 116222.
- Monaci, M., Agasucci, V., Grani, G. (2024). „An actor-critic algorithm with policy gradients to solve the job shop scheduling problem using deep double recurrent agents”, *European Journal of Operational Research*, 312(3), 910–926.
- Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., Kyek, A. (2018). “Optimization of global production scheduling with deep reinforcement learning”, *Procedia CIRP*, 72, 1264-1269.