Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

DIGITAL TWINS AND DEEP REINFORCEMENT LEARNING FOR ONLINE OPTIMIZATION OF SCHEDULING PROBLEMS

Bulent Soykan
Ghaith Rabadi

School of Modeling, Simulation, and Training
University of Central Florida
3100 Technology Parkway
Orlando, FL 32826, USA

ABSTRACT

This paper presents an approach that combines data-driven digital twins (DTs) and deep reinforcement learning (DRL) to address the challenges of online optimization of scheduling problems, focusing specifically on the classic job shop scheduling problem. Traditional approaches to solving such problems often encounter limitations in handling uncertainties and dynamic environments. In this study, we explore the integration of DTs and DRL to enhance decision-making in scheduling problems. We investigate the adaptability of a Graph Neural Network model within the DRL framework, enabling the agent to learn optimal scheduling policies through interactions with the DT. The potential of this convergence to tackle modern scheduling complexities offers insights into the future of operations management.

1 INTRODUCTION

Scheduling problems, including the classic Job Shop Scheduling Problem (JSSP), involve optimizing the allocation of resources to tasks over time. The JSSP is a well-known combinatorial optimization problem that requires determining the optimal schedule for processing n given jobs $(J = \{J_1, J_2, \ldots, J_n\})$ on m machines $(M = \{M_1, M_2, \ldots, M_m\})$ while adhering to specific constraints (sequence-dependent setup times etc.) and objectives (minimize the makespan, tardiness, maximum lateness etc.). Despite its name, JSSP has application areas beyond manufacturing and assembly lines. Due to its complexity and combinatorial nature, JSSP is a well-established problem in operations research and computer science literature and it has been an active area of research for decades. Traditional methods such as heuristic approaches, mathematical programming and metaheuristic methods have been used to solve this problem. However, as the landscape of real-world operations becomes increasingly dynamic and susceptible to uncertainties, these traditional methods experience limitations in addressing the challenges presented by the online optimization of the JSSP. The capacity to respond and adapt to uncertain events in real time has emerged as a critical characteristic of contemporary operations management. Despite recent increase in the importance of the online optimization of JSSP, it received relatively less attention. In essence, the online optimization of the JSSP necessitates a novel perspective that can harness the power of cutting-edge technologies such as data-based digital twins and model-based methodologies like deep reinforcement learning (DRL).

Digital twins (DTs) are widely considered as a transformational concept that integrates computational models with data-driven learning (Grieves and Vickers, 2017). DTs offer a powerful means to represent and simulate real-world systems, providing insights into different scenarios and disruptions. By creating digital counterparts of physical systems, DTs facilitate predictive analysis and decision-making. Integrating DT into online optimization can enable the exploration of various disruption scenarios and evaluate the robustness of scheduling solutions. Recent advancements in AI and machine learning, especially DRL, have shown promising results in various combinatorial optimization tasks. By learning from interactions with environments, DRL agents can adapt and make decisions in dynamic settings. The primary research questions addressed in this study are: To what extent can a DRL agent learn robust and resilient scheduling policies through trial-and-error interactions with the DT of JSSP? Does DT representation enable the exploration of various disruption scenarios, creating a rich training environment for the DRL agent?

2 DEEP REINFORCEMENT LEARNING AGENT AND THE DIGITAL TWIN

There are two primary categories of RL algorithms: policy-based and value-based. In policy-based approaches, the agent directly associates a state with a probabilistic distribution of potential actions, while value-based methods focus on estimating the anticipated cumulative reward for different actions, leading the agent to choose an action based on these predictions. The selection between these two approaches hinges on specific requirements and contextual factors. In policy-based RL, the agent's actions are determined by a policy $\pi(s_t, a_t)$, denoting the likelihood of taking action at in state s_t. Conversely, value-based methods take an indirect route, as they don't directly map states to actions. In the context of JSSP, choosing a policy-based approach is motivated by several factors. Policy-based methodologies provide a level of flexibility, adaptability, and the capacity to integrate domain expertise, rendering them highly suitable for tackling optimization challenges such as scheduling and routing.

DRL has emerged as a powerful paradigm for solving complex problems in various domains. However, due to the challenges of learning directly from the physical world, DRL often requires simulation environments for iterative learning through trial-and-error and obtaining optimal policies. One of the promising approaches in this regard is the utilization of DTs. DTs provide real-time data updates and contributes to improved decision-making and adaptability by providing a rich training environment for the DRL agent. At its core, DT is a dynamic simulation model driven by real-time data that is being updated continuously to represent the evolving real-world system accurately.

3 METHODOLOGY

The research methodology involves framing JSSP as a Markov decision process, where a scheduling agent interacts with the environment through a series of discrete decision points. Our environment is constructed based on the OpenAI Gym JSSEnv proposed by Tassel, Gebser, and Schekotihin (2021), with continuous updates derived from simulated data to represent real-time information. This augmentation introduces a rich source of diverse scenarios, including disruptions and fluctuations in job characteristics and machine availability. DRL agent is based on a Graph Neural Network (GNN) model that generates node embeddings from the graph-based state representation and learns the policy directly to select what action to take by calculating probability distribution over actions at that state. The model’s parameters are adjusted at each learning epoch to minimize the discrepancy between the predicted policies and the optimal policies generated in previous iterations. Notably, the GNN model can handle different numbers of nodes in the graph and still accurately calculate selection probabilities, even when the number of nodes change. This makes the DRL agent better at dealing with dynamic scheduling situations and uncertainties.

4 CONCLUSION

As DT technologies advance and DRL algorithms become more capable, the potential for enhanced decision-making in dynamic and uncertain environments becomes increasingly promising. By embracing this convergence, operations management can harness the power of digitalization and machine learning to tackle the challenges of modern scheduling problems.

REFERENCES

Grieves, M., & Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches* (pp. 85-113). Springer International.

Tassel, P., Gebser, M., & Schekotihin, K. (2021). A reinforcement learning environment for job-shop scheduling. *arXiv preprint arXiv:2104.03760*.