4th Conference on Production Systems and Logistics

Energy-Flexible Job-Shop Scheduling Using Deep Reinforcement Learning

Mine Felder[1], Daniel Steiner[1], Paul Busch[2], Martin Trat[1], Chenwei Sun[1], Janek Bender[1], Jivka Ovtcharova[1]

[1]*FZI Research Center for Information Technology, Karlsruhe, Germany*
[2]*Karlsruhe Institute of Technology, Karlsruhe, Germany*

Abstract

Considering its high energy demand, the manufacturing industry has grand potential for demand response studies to increase the use of clean energy while reducing its own electricity cost. Production scheduling, driven by smart demand response services, plays a major role in adjusting the manufacturing sector to the volatile electricity market. As a state-of-the-art method for scheduling problems, reinforcement learning has not yet been applied to the job-shop scheduling problem with demand response objectives. To address this gap, we conceptualize and implement deep reinforcement learning as a single-agent approach, combining electricity cost and makespan minimization objectives. We consider makespan as an ancillary objective in order not to entirely abandon the timely completion of production operations while assigning different weights to both objectives and analyzing the resulting trade-offs between them. Our main contribution is the integration of the electricity cost-related objective. We present two innovative reward functions, which consider the dynamic electricity prices to select a job for the machine or allow the machine idle. The reinforcement learning agent finds optimal schedules determined by cumulative electricity costs for benchmark scheduling cases from the literature.

Keywords

Deep Reinforcement Learning; Production Planning; Job-Shop Scheduling; Demand Response; Energy Flexibility

1. Introduction

Sustainable energy consumption has become a critical issue of the industry concerning its massive energy demand and related emissions. The industry sector is accountable for the largest share of electricity consumption with 42% in 2018 compared to other sectors including household, commerce and transportation [1]. According to [2], the global greenhouse emissions caused by industrial processes reached a record level in 2021 with 2.54 Gt CO_2eq. Demand response (DR) is one of the mechanisms to tackle the growing energy and emissions problem of the the industry, which has a key role in energy management. With the help of smart grid devices, DR aims to improve electricity grid reliability, increase the demand for renewable energy, and cut energy-related emissions. Furthermore, DR motivates end-users to plan their consumption based on time-varying electricity prices in order to reduce their electricity bills [3,4].

Regarding the planning of electricity consumption of manufacturing, researchers have developed several approaches for integrating the DR aspect into the scheduling and utilized various scheduling techniques. Through rapid developments in artificial intelligence and machine learning, reinforcement learning (RL) has recently brought new ground for scheduling problems [4,5]. By continuously learning through interaction

DOI: https://doi.org/10.15488/13454
ISSN: 2701-6277

with the problem's environment, RL can create optimal, generic and rapid solutions for dynamic and complex problems [3]. These features make RL a convenient method for handling the dynamic characteristics of manufacturing as well as electricity markets.

Although numerous studies on the application of RL in scheduling problems are available, only a few have considered energy management aspects [4,6]. None of these studies address job-shop scheduling (JSS) intending to minimize electricity costs. Therefore, in this study, we work on developing a deep RL (DRL) methodology for JSS with electricity cost and makespan minimization objectives. Although yet to be published, [7] show a highly promising RL approach for JSS to minimize makespan. This paper presents how we successfully augmented their approach with DR aspects. To do this, we develop two different reward functions for two machine states: operating and being idle. This study considers makespan minimization as an auxiliary objective only to ensure the completion of production operations on time, whereas the electricity cost objective is assessed more in-depth. We give weights to both objectives to analyze the trade-offs in between and provide flexibility in the decision-making process of production planners. Our results show that the RL agent can reduce electricity costs with the help of our two reward functions.

The rest of the paper continues as follows: Section 2 explains the background behind this study, including JSS, DR, the related literature and RL. Section 3 provides our data and RL model. Section 4 presents the results of the model and discussion. Finally, section 5 gives a summary and an outlook for the future work direction.

2. Background

2.1 Job-Shop Scheduling

Among various scheduling problems, JSS is characterized as a combinatorial optimization problem, aiming to find an optimal solution from a finite set of feasible ones. Although the number of possible solutions is finite, JSS is one of the hardest manufacturing problems to solve, being identified as NP-hard (non-deterministic polynomial-time hard) [7].

JSS has two main components, which are jobs and machines. Each machine performs a specific processing step of a job, called an operation, in a specific sequence. There exist precedence constraints between operations. Machines can only execute one operation at a time until it is finished [8].

2.2 Demand Response

Due to limited transmission and storage opportunities for electricity, electricity supply and demand should always be in balance [9]. To keep balance, DR motivates alterations on the demand side in order to compensate for dynamic changes in electricity supply owing to diverse factors, including weather for renewable energy sources. In addition to balancing supply and demand, another goal of DR is to increase the electricity demand for times with large generation from renewable sources [10].

Without emission costs, electricity generation from renewables has lower marginal costs than conventional energy sources [11]. This cost advantage also makes DR attractive for consumers to save on their electricity bills. As the most significant electricity consumption sector, the industry could potentially benefit from DR to a great extent. Many studies have shown the applicability and advantages of industrial DR [4]. Nevertheless, according to [3], participation in DR is far less in the industry sector than residential and commercial sectors due to several barriers. The modeling of DR should consider complex and interdependent industrial processes and various electricity consumption profiles depending on equipment and job [12]. Another issue is potential risks for daily production, such as losses or penalties because of the shift in schedule [13]. Thus, the manufacturing sector approaches DR cautiously. That is why this study considers the timely completion of production tasks while scheduling DR.

2.3 Related Work

There have been numerous studies on scheduling problems, focusing on diverse objectives. This section presents the literature only dedicated to energy and makespan-related objectives applied to various scheduling problems (not only JSS). The applied methods can be categorized as heuristics, mathematical models and RL.

Heuristics applied to scheduling problems by previous studies include simulated annealing algorithm [14], particle swarm optimization [15], colony optimization [16], backtracking search algorithm [17] and genetic algorithm [18]. Overall, heuristics are preferable for their practicality, but they do not ensure globally optimal solutions and mostly offer near-optimal solutions [7,19].

Researchers focus on mathematical models to find the globally optimal solution for scheduling problems, such as mixed integer linear programming [20] and constraint programming [21]. However, compared to heuristics, these methods require complex mathematical expressions, precise modeling for each specific scheduling problem and much longer computational times [4].

The application of RL to scheduling problems has received much attention recently and is already utilized in a variety of optimization and decision-making problems [5]. [8,19,22] perform various RL algorithms intending to minimize the makespan. Regarding energy-relevant objectives, the authors of [4] utilize multi-agent DRL for the flow-shop scheduling problem to minimize the cost of energy and materials for manufacturing, while [6] use single-agent RL for rescheduling the production after a machine breakdown to minimize makespan and electricity consumption. However, until now, there have not been many studies for energy-flexible JSS using RL. With this study, we aim to fill this gap in the literature.

2.4 Reinforcement Learning

RL is a branch of machine learning, and its naming originates from psychological reinforcement theory, which explains the ability to shape personal behaviors by reinforcement and punishment [23]. Figure 1 illustrates the fundamental components of RL and the interaction in between. The main components include an agent and the environment in that the agent is located. In RL, the agent learns how to achieve a goal by interacting with the environment.

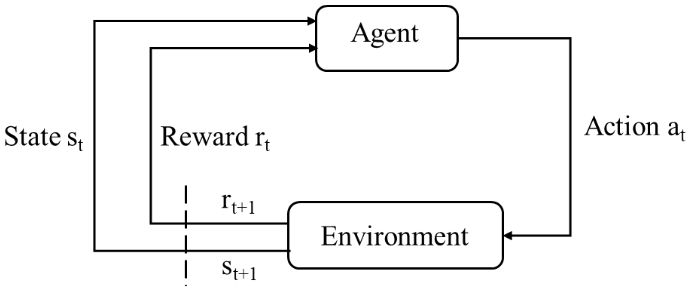

Figure 1: Representative deep reinforcement learning diagram (based on [4,24])

[24] states that the RL problem is based on an incompletely known Markov decision process where the agent partly observes the environmental state (s_t) and takes an action (a_t) regarding the observation. Taking an action (a_t) leads to a transition to the next state of the environment (s_{t+1}) and a numerical reward (r_{t+1}) from the environment. The RL agent tries to learn a policy that maps actions to the environmental states to maximize the numerical reward. The value function of the policy π is given by equation (1), which is the expected value (E) of discounted cumulative future rewards. In the equation, R stands for the set of rewards that the agent can obtain given that the state s at time t, while γ represents the discounting factor for future rewards [24].

$$V_\pi(s) = \mathrm{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | \, S_t = s\right] \tag{1}$$

A variety of RL algorithms have been developed so far, split into model-based and model-free algorithms. The main difference between the two categories is that model-based ones require modeling of state transition, whereas model-free ones do not. According to [5], due to the high computational effort for modeling state transition for a production environment, researchers preferred the model-free algorithms in scheduling problems more often, which are categorized as value-based and policy-based. Value-based algorithms try to learn state values V_π to find an optimal policy. In contrast, vanilla policy-based algorithms try to find an optimal policy by policy updates. Actor-critic methods, which are advanced policy-based algorithms, combine the advantages of these two methods by considering both the value function and policy update [25]. The main issue of actor-critic methods is to ensure a stable learning performance for the agent. As a DRL algorithm based on the actor-critic method, the proximal policy optimization (PPO) algorithm tries to ensure reliable learning performance by limiting policy updates with a hyperparameter called clipping parameter [26].

3. Methodology

We base our methodology on the work by [8], who build an RL environment to minimize makespan in the JSS problem. The following subchapters present our modifications on this JSS environment and the benchmark JSS datasets from the literature in order to integrate our primary objective of electricity cost minimization and to provide flexibility in production planning.

3.1 Data preparation

We implement our modified JSS model on the first ten instance datasets from [27]. The datasets are identical in size, including 15 machines and 15 jobs, each job has 15 operations. [27] gives processing times and machine numbers as the statistical data to describe an operation.

Our work in this stage includes the integration of external electricity data into the instance datasets. As external data, we add electricity demand per operation (kW) and the German day-ahead electricity prices (€/kWh). Electricity demand data is sampled from a discrete uniform distribution over the unit interval [1, 20] and given for each operation. We obtain the German day-ahead electricity prices from EPEX Spot through the Price API developed by aWATTar [28]. The data includes the prices between 11.10.2022 at 15:00 and 13.10.2022 at 00:00.

3.2 Modeling

In this stage, we enhance the JSS environment by considering energy-flexible scheduling. Specifically, the major modifications occur in the observation space and reward function, whereas the action space stays the same as the original version.

3.2.1 Action Space

The authors of [8] design the action space as the selection of a job for machines or keeping them idle, called "No-Op" by the authors. It is defined by the discrete space with the input $\{j_0, j_1, \ldots, j_{J-1}, No\text{-}Op\}$. A boolean vector, representing the legal actions, eliminates ineligible actions, such as selecting finished or ongoing jobs or occupied machines, which is explained in Section 3.2.2. Additionally, the authors implement the "action masking" technique that equates the likelihood of taking illegal actions close to 0%. We employ these action-related settings without an alteration since our electricity cost objective does not require any change in the action space.

3.2.2 Observation Space

In the JSS environment from [7], the observation space of the agent includes seven attributes about required and leftover times for operations as well as idle times. Regarding our objective of electricity cost minimization, we add 2 new electricity-related attributes to the agent's observation space: the electricity demand of the operation and weighted electricity price. We weigh the hourly electricity price by the processing time because operations might take longer than one hour. The authors scale the majority of the attributes by the associated maximal values to treat them equally and benefit from the stability of the gradient calculation of the PPO network. We scale our two attributes accordingly. Table 1 shows all attributes of the observation space with explanations.

Table 1: Observation space attributes

State attribute	Description	Scaling Parameter	Range
s_1	Boolean parameter shows if jobs can be allocated to the machine	-	{0, 1}
s_2	Left over time for the currently performed operation	Longest operation time	[0,1]
s_3	Percentage of finished operations of a job	-	[0,1]
s_4	Left over time until total completion of a job	Longest job completion time	[0,1]
s_5	Required time until the machine is free	Longest operation time	[0,1]
s_6	Idle time since last job's performed operation	Durations of all operations	[0,1)
s_7	Cumulative idle time for the job in the schedule	Durations of all operations	[0,1)
s_8	Electricity demand of the operation	Largest electricity demand	[0,1]
s_9	Weighted electricity price	Highest electricity price	[0,1]

3.2.3 Reward Function

Equation (2) shows the original reward function from [8], which is an approach to minimize the makespan, calculated by taking the difference between the processing time of operation *i* of job *j* (p_{ij}), and idle times on machine *m* ($empty_m$), caused by selecting job *j* between old (*s*) and new states (*s'*). The authors scale the reward *R(s,a)* by maximum operation length.

$$\boldsymbol{R(s,a) = p_{ij} - \sum_{m\in M} empty_m(s,s')} \tag{2}$$

We develop two separate reward functions for selecting a job as the action and the "No-Op" action. Equation (3) represents the first reward function for the action for selecting a job. α represents the weight parameter for two objectives: makespan and electricity cost minimization. We integrate the electricity costs at the end of the formula. e_{ij} refers to the electricity demand of operation *i* of job *j* and π^h stands for the electricity price at hour *h*. p_{ij}^h represents the fraction of the processing time, which falls into each pricing hour. We scale the electricity costs by the largest electricity cost possible, which is the product of the highest electricity price and demand in the datasets. We subtract the weighted electricity costs from one, because the reward and electricity cost are inversely related. The agent receives less reward if it chooses an energy-intensive operation during an hour with a high electricity price.

$$R_1(s,a) = (1-\alpha)\left(p_{ij} - \sum_{m\in M} empty_m(s,s')\right) + \alpha\left(1 - \sum e_{ij}\,\pi^h p_{ij}^h\right) \tag{3}$$

For the “No-Op” action, the reward function given in equation (4) cannot have the operation-related parameters (e_{ij} and p_{ij}^{h}). Instead, we add only the electricity price to the reward function and scale by its maximum value. There is a different concept behind the reward function in equation (4) than in equation (3). On the one hand, the idle operation is not favorable for makespan minimization, because it can extend the total makespan. On the other hand, when the prices are high, it can be advantageous to not operate the machine to reduce the electricity cost. This concept is reflected in the change of the sign before the weight of the electricity minimization objective.

$$R_2(s,a) = (1-\alpha)\left(-\sum_{m\in M} empty_m(s,s')\right) - \alpha\left(1-\sum \pi^h\right) \qquad (4)$$

4. Results and Discussion

We execute the PPO algorithm on a computer with AMD Ryzen 9 3950X CPU and 2 Nvidia Titan RTX GPU. We employ the same training time (10 minutes) and hyperparameters as [8].

Table 2 presents the results with all instance datasets. The results clearly show the antagonistic effect of α on the makespan and the electricity cost. Larger α values result in lower electricity costs and longer makespan. There are a few exceptions to this, such as the electricity costs of the datasets Taillard-03, Taillard-04 and Taillard-06. When we increase α from 0.5 to 0.8 for these datasets, we obtain higher electricity costs. In addition, incrementing α from zero to 0.2 provides shorter makespan values for the datasets Taillard-01 and between Taillard-05 and Taillard-09.

We obtain the significant difference, once we set α to 1, giving no focus on the makespan objective. Our reward functions provide an average 13% reduction in electricity costs when we focus fully on the electricity cost objective (α=1) compared to the model with full focus on makespan (α=0). Among the datasets, Taillard-09 has the biggest electricity cost savings with 18%. However, this rate may also depend on the electricity price data used for electricity cost calculation. The range of our price data lies between 0.2 and 0.49 €/kWh.

Table 2: Results in terms of electricity costs and makespan for the instance datasets with different weights

	Electricity costs (Euro)					Makespan (minute)				
Dataset	α=0	α=0.2	α=0.5	α=0.8	α=1	α=0	α=0.2	α=0.5	α=0.8	α=1
Taillard-01	687.20	671.63	659.08	651.38	574.37	1335.03	1310.48	1331.00	1343.05	1978.59
Taillard-02	589.64	572.66	572.15	539.43	497.22	1323.00	1323.00	1295.55	1421.28	1813.57
Taillard-03	596.35	595.96	581.99	583.99	546.44	1317.04	1333.01	1339.45	1338.11	1851.42
Taillard-04	608.01	597.18	579.19	582.67	536.16	1226.49	1264.63	1291.00	1333.80	1863.42
Taillard-05	587.55	574.10	556.34	549.01	507.46	1339.00	1309.04	1400.99	1427.38	1715.81
Taillard-06	569.59	546.69	545.10	553.94	495.90	1347.26	1325.25	1331.00	1366.76	1927.81
Taillard-07	608.94	620.05	607.67	600.00	566.02	1342.73	1294.12	1346.53	1412.49	1781.08
Taillard-08	598.03	599.66	596.70	594.88	497.15	1324.37	1319.08	1319.14	1349.10	2093.88
Taillard-09	620.32	594.84	597.37	581.84	511.03	1424.19	1392.53	1409.62	1506.72	2085.78
Taillard-10	609.57	598.08	596.76	581.88	528.85	1320.68	1367.03	1397.00	1413.77	1939.35

Figure 2 illustrates the evolution of makespan and electricity cost for the last instance data (Taillard-10) during the training with regard to α. At the beginning of the training, electricity costs are low since the agent has still jobs to finish. The costs rise as the agent assigns these jobs to machines. After reaching a peak, the agent is able to reduce the costs continuously through the half of the training. The results are stable afterwards. For α=1, the electricity cost is the lowest with €528.85 and the makespan is the longest with 1939.35 min at the end of the training. Contrarily, for α=0, the electricity cost is the largest with €609.57, while the makespan is the shortest with 1320.68 min. The intermediate α values lead to sequential results between these peak values. However, their results are closer to the results with α=0 rather than α=1. This may indicate that the makespan objective has a greater influence on the agent than the electricity cost objective for intermediate α values.

Figure 3 compares the predictive schedules for the same instance data (Taillard-10) with the maximum and minimum α values. The left schedule focuses on makespan minimization and shows that the agent is able to reduce the idle times of machines by limiting the "No-Op" action. The right schedule focuses completely on the electricity cost objective and illustrates how the agent distributes the operations over the longer time horizon by considering the electricity prices and demands of the operations. The agent completes the majority of the operations during the first 21 hours in the right schedule, similar to the left schedule. Three out of 15 jobs (Job 0, 6 and 7) are postponed to hours between 21 and 33. Completing 80% of the jobs within the same time as in the left schedule, the agent is still able to reduce the electricity costs from €609.57 to €528.85 in the right schedule (see Table 2).

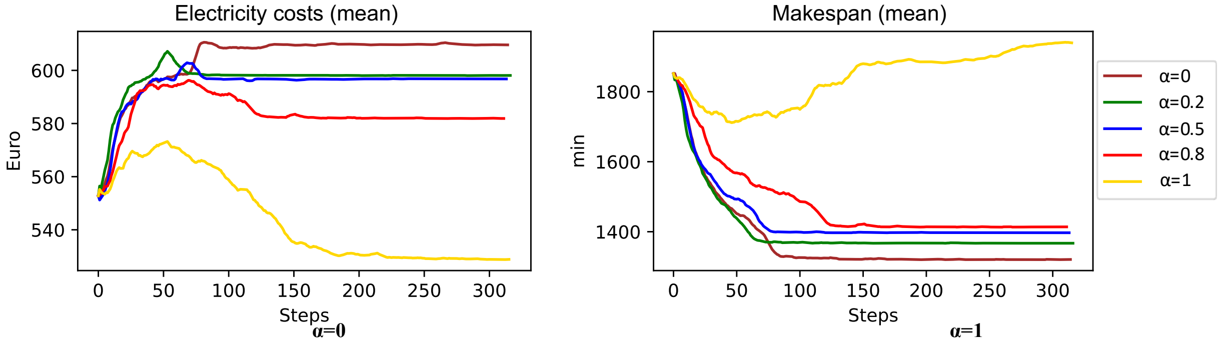

Figure 2: Evolution of the mean electricity costs and makespan during the training

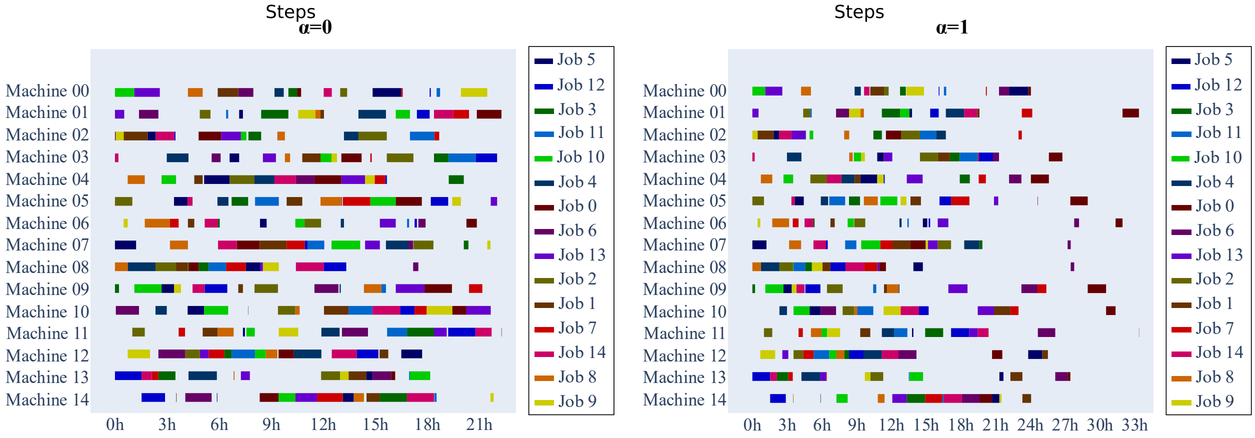

Figure 3: The estimated schedules with full weights on makespan (α=0) and electricity costs objectives (α=1)

5. Summary and Outlook

In this study, we aim to present the RL-based method for energy-flexible production planning on the job-shop problem considering the timely completion of production tasks. For this purpose, we integrate the objective of electricity cost minimization into the RL environment built by [8] for makespan minimization. Our key contribution is the development of two distinct reward functions for operating a job and leaving the machine idle. Using the weight parameter for the objectives of makespan and electricity cost minimization, the algorithm provides flexibility in production planning, considering the preferences of the manufacturing company between the makespan of jobs and electricity costs. The results clearly show the trade-off situation between two objectives. Implementation on the benchmark scheduling cases show that giving full weight to the electricity cost objective can reduce electricity costs by 13% on average compared to giving full weight to makespan objective. One of the benchmark cases indicates that the agent schedules 20% of the jobs to a later time to decrease the electricity costs, while 80% are completed within the optimal time determined by the algorithm with full weight given to minimize the makespan.

Future research work may build on our findings by testing our agent for more available instance data and comparing PPO against other RL algorithms. A significant research direction would be adapting the algorithm to the more complicated form of JSS: the flexible JSS, which includes many available machines for each operation and different electricity demands for machine-operation pairs.

Acknowledgements

We declare that this work was funded by the German Federal Ministry for Economic Affairs and Climate Action (grant number: 03EI4018C, Delfine project).

References

[1] International Energy Agency (IEA). Key World Energy Statistics 2020.

[2] IEA, International Energy Agency, 2022. Global Energy Review: CO2 Emissions in 2021. https://www.iea.org/reports/global-energy-review-co2-emissions-in-2021-2.

[3] Huang, X., Hong, S.H., Yu, M., Ding, Y., Jiang, J., 2019. Demand Response Management for Industrial Facilities: A Deep Reinforcement Learning Approach. IEEE Access 7, 82194–82205.

[4] Lu, R., Li, Y.-C., Li, Y., Jiang, J., Ding, Y., 2020. Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management. Applied Energy 276, 115473.

[5] Wang, L., Pan, Z., Wang, J., 2021. A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling. Complex Syst. Model. Simul. 1 (4), 257–270.

[6] Naimi, R., Nouiri, M., Cardin, O., 2021. A Q-Learning Rescheduling Approach to the Flexible Job Shop Problem Combining Energy and Productivity Objectives. Sustainability 13 (23), 13016.

[7] Chen, R., Yang, B., Li, S., Wang, S., 2020. A self-learning genetic algorithm based on reinforcement learning for flexible job-shop scheduling problem. Computers & Industrial Engineering 149, 106778.

[8] Tassel, P., Gebser, M., Schekotihin, K., 2021. A Reinforcement Learning Environment For Job-Shop Scheduling, 7 pp. http://arxiv.org/pdf/2104.03760v1.

[9] van der Veen, R.A., Hakvoort, R.A., 2016. The electricity balancing market: Exploring the design challenge. Utilities Policy 43, 186–194.

[10] Müller, T., Möst, D., 2018. Demand Response Potential: Available when Needed? Energy Policy 115, 181–198.

[11] Do, L.P.C., Lyócsa, Š., Molnár, P., 2019. Impact of wind and solar production on electricity prices: Quantile regression approach. Journal of the Operational Research Society 70 (10), 1752–1768.

[12] May, G., Stahl, B., Taisch, M., 2016. Energy management in manufacturing: Toward eco-factories of the future – A focus group study. Applied Energy 164, 628–638.

[13] McKane, A.T., Piette, M.A., Faulkner, D., Ghatikar, G., Radspieler Jr., A., Adesola, B., Murtishaw, S., Kiliccote, S., 2008. Opportunities, Barriers and Actions for Industrial Demand Response in California, 89 pp.

[14] Keller, F., Schultz, C., Braunreuther, S., Reinhart, G., 2016. Enabling Energy-Flexibility of Manufacturing Systems through New Approaches within Production Planning and Control. Procedia CIRP 57, 752–757.

[15] Dababneh, F., Li, L., Shah, R., Haefke, C., 2018. Demand Response-Driven Production and Maintenance Decision-Making for Cost-Effective Manufacturing. Journal of Manufacturing Science and Engineering 140 (6).

[16] Jia, Z., Wang, Y., Wu, C., Yang, Y., Zhang, X., Chen, H., 2019. Multi-objective energy-aware batch scheduling using ant colony optimization algorithm. Computers & Industrial Engineering 131, 41–56.

[17] Caldeira, R.H., Gnanavelbabu, A., Vaidyanathan, T., 2020. An effective backtracking search algorithm for multi-objective flexible job shop scheduling considering new job arrivals and energy consumption. Computers & Industrial Engineering 149, 106863.

[18] Wei, H., Li, S., Quan, H., Liu, D., Rao, S., Li, C., Hu, J., 2021. Unified Multi-Objective Genetic Algorithm for Energy Efficient Job Shop Scheduling. IEEE Access 9, 54542–54557.

[19] Bouazza, W., Sallez, Y., Beldjilali, B., 2017. A distributed approach solving partially flexible job-shop scheduling problem with a Q-learning effect. IFAC-PapersOnLine 50 (1), 15890–15895.

[20] Fang, K., Uhan, N., Zhao, F., Sutherland, J.W., 2011. A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction. Journal of Manufacturing Systems 30 (4), 234–240.

[21] Raileanu, S., Anton, F., Iatan, A., Borangiu, T., Anton, S., Morariu, O., 2017. Resource scheduling based on energy consumption for sustainable manufacturing. J Intell Manuf 28 (7), 1519–1530.

[22] Samsonov, V., Kemmerling, M., Paegert, M., Lütticke, D., Sauermann, F., Gützlaff, A., Schuh, G., Meisen, T., 2021. Manufacturing Control in Job Shop Environments with Reinforcement Learning, in: Proceedings of the 13th International Conference on Agents and Artificial Intelligence. Science and Technology Publications, pp. 589–597.

[23] Ernst, D., Glavic, M., Capitanescu, F., Wehenkel, L., 2009. Reinforcement learning versus model predictive control: a comparison on a power system problem. IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society 39 (2), 517–529.

[24] Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. Second edition. The MIT Press, Cambridge, Massachusetts, 352 pp.

[25] Konda, V. R.V., Tsitsiklis, J.N., 1999. Actor-Critic Algorithms.

[26] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal Policy Optimization Algorithms, 12 pp. http://arxiv.org/pdf/1707.06347v2.

[27] Taillard, E., 1993. Benchmarks for basic scheduling problems. European Journal of Operational Research 64 (2), 278–285.

[28] aWATTar. API - Preis Datenfeed. https://www.awattar.de/services/api. Accessed 1 October 2022.

Biography

Mine Felder studied Industrial Engineering (B.Sc.) at the Izmir University of Economics and Sustainable Resources Management (M.Sc.) at the Technical University of Munich. She is working as a research scientist in the research unit Intelligent Systems and Production Engineering (ISPE) at the FZI Research Center for Information Technology. Her research focuses on sustainable and smart manufacturing using machine learning methods.

Daniel Steiner is studying computer science (M.Sc.) at the Karlsruhe Institute of Technology (KIT), majoring in anthropomatics and cognitive systems as well as embedded system design and computer architecture. Since 2021, he works as a student assistant at FZI Research Center for Information Technology in Karlsruhe with a focus on machine learning in manufacturing.

Paul Busch is studying Mechanical Engineering (M.Sc.) at the Karlsruhe Institute of Technology (KIT). In his graduate thesis, he is working on the topic of the energy-flexible job shop scheduling in cooperation with the FZI Research Center for Information Technology and the KIT.

Martin Trat studied at the Karlsruhe Institute of Technology, where he majored in Industrial Engineering. He works as a research scientist at the FZI Research Center for Information Technology in Karlsruhe and focuses his research activities on the efficient and robust application of productive artificial intelligence in dynamically changing environments.

Chenwei Sun studied automotive engineering (B.Sc.) at Hefei University of Technology and mechanical engineering (M.Sc.) at Karlsruhe Institute of Technology (KIT), majoring in automotive engineering and information technology. Since 2020, he has been working as a research assistant at the FZI Research Center for Information Technology in Karlsruhe.

Janek Bender is vice department manager in the research unit Intelligent Systems and Production Engineering (ISPE) at the FZI Research Center for Information Technology in Karlsruhe. His research interests revolve around applying artificial intelligence to the manufacturing domain, especially within the context of job scheduling.

Jivka Ovtcharova is a full professor, head of the Institute for Information Management in Engineering (IMI) since October 2003 and founder of the Lifecycle Engineering Solutions Center (LESC) at the Karlsruhe Institute of Technology (KIT), established in June 2008. In addition, she is the director for Process and Data Management in Engineering (PDE) at the FZI Research Center for Information Technology in Karlsruhe since 2004.