2nd Conference on Production Systems and Logistics

Deep Reinforcement Learning In Production Planning And Control: A Systematic Literature Review

Marcel Panzer[1], Benedict Bender[1], Norbert Gronau[1]
[1]*Chair of Business Informatics, esp. Processes and Systems, University of Potsdam, Potsdam, Germany*

Abstract

Increasingly fast development cycles and individualized products pose major challenges for today's smart production systems in times of industry 4.0. The systems must be flexible and continuously adapt to changing conditions while still guaranteeing high throughputs and robustness against external disruptions. Deep reinforcement learning (RL) algorithms, which already reached impressive success with Google DeepMind's AlphaGo, are increasingly transferred to production systems to meet related requirements. Unlike supervised and unsupervised machine learning techniques, deep RL algorithms learn based on recently collected sensor- and process-data in direct interaction with the environment and are able to perform decisions in real-time. As such, deep RL algorithms seem promising given their potential to provide decision support in complex environments, as production systems, and simultaneously adapt to changing circumstances.

While different use-cases for deep RL emerged, a structured overview and integration of findings on their application are missing. To address this gap, this contribution provides a systematic literature review of existing deep RL applications in the field of production planning and control as well as production logistics. From a performance perspective, it became evident that deep RL can beat heuristics significantly in their overall performance and provides superior solutions to various industrial use-cases. Nevertheless, safety and reliability concerns must be overcome before the widespread use of deep RL is possible which presumes more intensive testing of deep RL in real world applications besides the already ongoing intensive simulations.

Keywords

Deep Reinforcement Learning; Machine Learning; Production Planning; Production Control; Systematic Literature Review

1. Introduction

Today's production has to cope with significantly increased complexities due to accelerating innovation cycles and increasingly individualized customer demands. Fully customized products and on-demand production impose high challenges on the associated production systems. In particular, production planning and control must be able to deal with uncertainties and constantly changing production environments [1]. Failures must be compensated quickly to enable on-time deliveries and optimize the production performance. Besides, production logistics must be able to perform the planned actions and meet the same requirements to ensure high robustness and reduce downtimes [2].

One opportunity to fulfill the demanding requirements and to keep up with product development is the application of machine learning in production systems such as (semi-)supervised, unsupervised, or reinforcement learning (RL). In contrast to (semi-)supervised and unsupervised learning, RL does not require a prelabeled set of data and any human supervision. It is characterized in particular by its trial-and-error learning

approach in direct interaction with the environment [3] and enables real-time online decision-making and an adaptive system design [4]. Especially with the success of DeepMind´s AlphaZero [5], neural network based RL received special attention which has resulted in a large number of publications in various fields and emphasized its capabilities in complex systems. However, even though [6] already emphasized the potential of general machine learning in production to improve quality and increase performances and availabilities, no focused review on research outcomes was conducted for the deployment of deep RL in production in recent years. In contrast, the fields of CPS [7] or general economics [8] among others have outlined research findings in a bundled manner and elaborated the advantages as well as major issues yet to be solved.

We intend to provide a systematic literature review of ongoing deep RL research in production planning, control, and logistics. This includes the identification of simulated and real-world implementations as well as current implementation challenges. We also want to derive possible future research directions and provide incentives to leverage the deployment of deep RL in applications that can benefit from its flexibility and adaptability. For this purpose, we intend to answer the following research questions in production planning, control and logistics:

- RQ1: What deep RL applications exist in the field of production planning, control and logistics?
- RQ2: What are existing implementation challenges of deep RL?
- RQ3: What are future research fields that need to be addressed to overcome these challenges and support implementations of deep RL in production systems?

To answer these research questions, we first give a short introduction to deep RL in Section 2, followed by the applied review methodology in Section 3. Section 4 presents the results of our review analysis (RQ1). Section 5 addresses RQ2 by outlining existing challenges and RQ3 by giving incentives for potential future research. Finally, a conclusion is given in Section 6.

2. State-of-the-art

RL is based on the agent-environment interaction loop as illustrated in Figure 1 and can be described as a sequential decision-making process. The agent performs an action and receives in turn a reward for this action and the current environmental state. With each loop and the gathered experience, the agent can adapt its behavior policy accordingly [3].

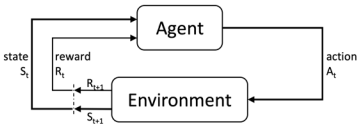

Figure 1: Agent-environment interaction loop [3]

Conventional RL methods often employ a Q-table for mapping the policy, in which recommendations for given states and the resulting actions can be retrieved. However, in high dimensional problem spaces, this leads to the curse of dimensionality and declining performances [9]. To circumvent this limitation, a neural network can be employed to map the policy. First demonstrated in 2013, such a deep RL algorithm outperformed human benchmarks in performances within the Atari environment [10]. Apart from the neural network, it is essential to distinguish between model-free and model-based algorithms. Model-based algorithms learn a general model of their environment and can make predictions about the next possible state. Model-free algorithms, on the other hand, do not learn a model of their environment, but iteratively gather experience and exploit their policy to evaluate executable actions [3]. Model-free algorithms can further be classified into value-based algorithms that require a discretization of the action space but have a better sample efficiency like the DQN, policy-based algorithms like a PPO that learn the policy directly and don´t need to evaluate actions based on Q-values like the DQN, and hybrid algorithms that try to combine both previously mentioned methods benefits [11].

A further characteristic of deep RL is its suitability for an application in decentralized and distributed multi-agent systems. Especially in the field of smart manufacturing and Industry 4.0, distributed systems can leverage the adaptability of a system and enable more robust responses against uncertainties and unforeseen events [12]. This makes RL being a promising technique to improve the performance of modern production systems.

3. Research methology

Before conducting the analysis, it is essential to establish a systematic review procedure to ensure a representative coverage of results for deep RL applications in production systems. In the further course, we follow the guidelines proposed by [13] and [14] and focus on the taxonomy as outlined in Table 1.

Table 1: Taxonomy framework

Characteristic	**Categories**			
Focus	Research outcomes	Research methods	Theories	Applications
Goal	Integration	Criticism	Central issues	
Perspective	Neutral representation	Espousal of position		
Coverage	Exhaustive	Ex. and selective	Representative	Central/pivotal
Organization	Historical	Conceptual	Methodological	
Audience	Specialized scholars	General scholars	Practitioners	General public

During the review process, we focus on research outcomes and applications of deep RL in production systems. Thereby we try to give integrative insights but also maintain a neutral position to highlight central issues that may block an implementation but also serve as research opportunities. Within our review scope, we provide a representative coverage of our chosen topic that addresses practitioners and general scholars.

For further refinement, we defined the keywords as listed in Table 2. Besides an artificial intelligence subset, a second subset describes the production domain, and a third the respective discipline.

Table 2: Defined keyword combinations

Deep RL subset			Domain subset		Discipline subset
Deep reinforcement learning OR Reinforcement learning AND	Artificial intelligence OR Deep learning OR Machine learning	AND	Production OR Manufacturing OR Assembly OR Automation	AND	Planning OR Control OR Scheduling OR Dispatching OR Logistics

During the review, we screened the retrieved literature from Web of Science, IEEE Xplore, and ScienceDirect (Title, abstracts and keywords, similar to [15]) according to pre-defined inclusion and exclusion criteria. We only considered English-language papers published after 2010, as deep RL and particular achievements in this field were achieved after the publication of [10] in 2013. In addition, only papers that received a peer review were included to ensure a high review quality. We included papers that focus on the impact of deep RL in production planning, control, and logistics. Purely technical papers or papers that focus on the development of algorithms were excluded. A summary of the review process is given in Figure 2. Remarkably, a large number of publications was excluded after the full-text review, since RL was considered as a machine learning technique, but did not utilize a neural network for the task completion.

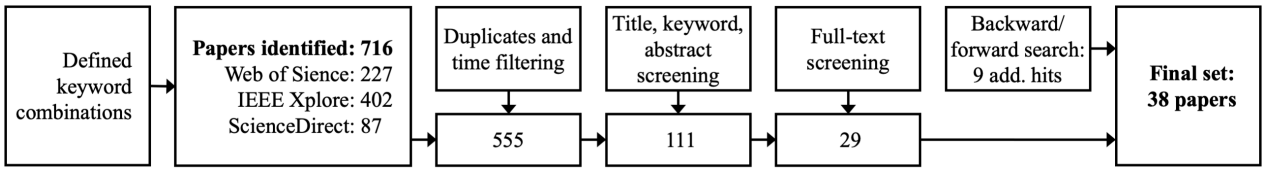

Figure 2: Review process

The distribution of publications over the years is illustrated in Figure 3. It is noticeable that no papers were published until 2017, but since 2018 there has been a significant increase, which underlines the ongoing focus on deep RL applications in current production research. On the other hand, the distribution among journals with more than one publication in Figure 4 indicates the high quality and relevance of published papers that appeared in highly recognized outlets.

Figures 3/4: Yearly publications and outlet contributions (2021 up to February)

4. Review analysis

During the in-depth analysis of the reviewed papers, production planning, as the discipline with the most publications (21), is considered first. Afterwards, due to the high overlap, production control (11) and logistics (6) are addressed in detail.

4.1 Production planning

In the field of production planning and especially scheduling, most of the reviewed papers were driven by the need to find more robust solutions that can deal with planning uncertainties and unforeseen incidents. Conventional algorithms have difficulties to cope with the dynamic production environment and often rely on human intervention or experience [16].

To deal with uncertainties, [17] increased a chemical plant's profitability by applying a policy-based algorithm for chemical production scheduling, outperforming the otherwise commonly employed MILP algorithm. Similarly, conventional algorithms in this field often have difficulties in dealing with down-times, delays, and rush orders, whereas deep RL algorithms demonstrate increased robustness. This is also evident in [18], who leveraged production scheduling to cope with highly spontaneous orders in the medical mask production during times of Covid-19. Additionally, fitted with a long-short term network, the algorithm responded more flexibly to inputs and operated faster with a reduced tardiness. Another set of papers focused on job-shop scheduling and reduced the makespan compared to FIFO, LPT/SPT, or other heuristics [19][20][16] or optimized tardiness levels, profits, and utilization rates [21]. To be more adaptive to differing problem granularities, [22] decomposed the general objective into a local and a global optimization and training problem.

In general, it is noticeable that 70% of the papers in the field of production scheduling utilized value-based algorithms that require a discretization of the action space. However, this is often feasible and can correspond to the selection of defined operations to reduce process time [23], or the selection to allocate a product to a specific machine [24]. On the other hand, a variety of inputs can be processed as demonstrated in [25] and

[26], who take Gantt diagrams for rescheduling processes as inputs. This not only led to a reduced tardiness, but also increased the flexibility of the system in handling diminishing shop-floor predictabilities.

Due to the great flexibility of the algorithm's reward function, it can be flexibly adapted and trained for other production optimization problems. To provide a brief application overview, the following Table 3 consolidates the reviewed papers within the scope of production scheduling. Even though it becomes evident that almost all of the proposed solutions outperformed conventional algorithms, they were all evaluated in simulations.

Table 3: Deep RL applications in production scheduling

Schedulings process	Superiority to conv. methods	Objective	Source
Chemical prod. scheduling	Superior	Maximize profits	[17]
Cloud manufacturing	Superior	Maximize utilization	[27]
Dynamic scheduling	-	Minimize completion time	[28]
	Superior	Minimize makespan	[29]
Flow shop scheduling	Superior	Reduce tardiness	[18]
Job-shop scheduling	Superior	Minimize makespan	[30]
	Superior	Minimize makespan	[19]
	-	Minimize makespan	[22]
	-	Minimize processing time	[31]
Job-shop scheduling	Superior	Minimize utilization and profits	[21]
	Superior	Minimize makespan	[20]
	Superior	Minimize makespan	[16]
Lot scheduling	Superior	Min. waiting, impr. cost-rates	[24]
Mold scheduling	Superior	Minimize processing time	[23]
Multichip production	Superior	Minimize makespan	[32]
Packaging line scheduling	Superior	Min. comp. time, energy cons.	[33]
Paint job scheduling	Superior	Minimize change-over costs	[34]
Parallel, re-entrant prod.	Comparable	High short-term return	[35]
Rescheduling	Lower tardiness	Reduce tardiness	[25]
	Lower tardiness	Reduce tardiness	[26]
Single machine scheduling	Superior	Minimize makespan, lateness	[36]

4.2 Production control

In production control, a key challenge is to compensate for sharp fluctuations in demand and breakdowns that occur at short notice to dispatch products to their respective target machines. Depending on the current state of production, orders have to be allocated to eligible machines according to their capacity, buffer levels, and further factors, while optimizing both local and global objectives [37]. To cope with the existing uncertainties, conventional methods require high computational efforts to adapt to process variations [38] or rely on single methods that do not operate optimally in each situation [39]. The static sequencing rule problem was addressed by [40] and [39], based on a situational sequencing rule selection. According to the current occupancies, machine status and others, the deep RL algorithm selected the best dispatching heuristic for the current production (such as FIFO) and significantly improved tardiness in most cases. Further approaches of adaptive job-shop scheduling were particularly investigated within the highly volatile and technically demanding wafer fabrication. By implementing deep RL driven dispatching rules, superior performances were reached compared to a variety of conventional methods, resulting in minimized time constraint violations and maintained WIP levels [41], increased machine utilization and reduced lead times [42], as well as simultaneously minimized utilization, throughput and waiting times [43]. Another approach to reduce WIP levels was proposed by [44] in production flow control. Compared to a maximum throughput method, the average WIP level could thereby be reduced by 43% with a minimal decrease in throughput (-0.2%). Further applications are listed in Table 4 and include a short-term decision-making process in mineral processing [4]

which increased cumulative cash flow by 15%, a multi-agent dispatching to optimize delivery performance within the semi-conductor industry [45], and a general transfer learning supported deep RL approach in job shop processes [46]. As in scheduling, most approaches (7 out of 11) outperformed conventional algorithms, but were again solely implemented in simulations.

Table 4: Deep RL applications in production dispatching

Dispatching process	Superiority to conv. methods	Objective	Source
General job-shop disp.	Comparable	Global and local optimization	[38]
	Comparable	Minimize mean tardiness	[39]
	Superior	Minimize total tardiness	[40]
	Superior	Minimize mean lateness/tardiness	[46]
Short-term mineral flow	Superior	Optimize profits, min. target deviations	[4]
Semiconductor	Comparable	Optimize delivery performance	[45]
	Comparable	Global and local optimization	[37]
Wafer fabrication	Superior	Optimize util., TH/waiting times	[43]
	Superior	Optimize util., lead times	[42]
	Superior	Min. time constr. violations, WIP	[41]
WIP bounding	Reduces WIP	Opt. through-put and WIP trade-off	[44]

4.3 Production logistics

In [47], a real-time intralogistics solution was proposed to handle uncertainties with autonomous mobile robots (AMR). Based on the states of the individual agents, they could negotiate orders and virtually raised bids which outperformed conventional methods in terms of logistics efficiency. In a similar scenario the deep RL algorithm determined optimal target machines for the automated guided vehicles (AGV) based on job information, queue sizes, and station status which reduced lead times compared to conventional methods [48]. For the orchestration of AGVs [49] implemented a mixed rule approach. Compared to single heuristics, the makespan and delay ratio were thus reduced by approximately 10%. Besides automated vehicles, [50] proposed a deep RL algorithm for the 3-grid sorting system control to fasten up product dispatching and enable multiple sorting objectives. Other applications were a collaborative robot conveyor belt processing to fill surrounding trays [51] and a syringe filling or virtual commissioning process, among others, which outperformed human benchmarks [52]. While all compared approaches were again able to improve benchmark performances, 5 out of 6 were evaluated in simulated environments.

Table 5: Deep RL applications in production logistics

Dispatching process	Superiority to conv. methods	Objective	Source
AGV control	Superior	Min. makespan and delay ratios	[49]
	Superior	Optimize lead-times	[48]
AMR control	Superior	On-time order completion	[47]
Robot batching	-	Reach target weights and opt. filling	[51]
Syringe filling process	Above human	Min. interruptions and bad decisions	[52]
Three-grid sorting system	-	Optimize in-/outflow control	[50]

5. Implementation challenges and potential research opportunities

Despite the superior performance of deep RL, we identified the algorithm and parameter selection and optimization as well as the simulation to reality transfer as major challenges that have to be overcome to fully leverage its potential in production systems.

Beginning with the optimization, there is no guarantee for an optimal solution [17] and it is necessary to consider local as well as global optimization measures, to prevent sub-optimal solutions and decreased performances [37, 38, 45]. Besides, the choice of the algorithm, the network parameters, and further adjustment possibilities must be clarified before implementation. Regarding the algorithm selection, 22 out of 26 value-based implementations utilized a DQN, which was often inferior to enhanced versions such as a dueling or double DQN [53] and may negatively impact the performance in the particular use-case. Further challenges arise from obtaining the desired reliability and safety of the proposed solution. In production, seamless operation without incidents and maximum predictability of the system must be constantly ensured. This can only be realized by transferring the results from the simulations to reality and through subsequent intensive in-process validations. However, such a transfer to a real production system was often considered critical or required great efforts, resulting in only a few conducted real-world testings and appropriate conclusions for reality are rather hard to derive.

One way to address the above mentioned challenges in future research is, first, to test and optimize similar algorithms in parallel, which can be implemented without much efforts and contribute to a more based performance testimonial. Second, the choice of the neural network parameters can be intensively adjusted beforehand to exploit the algorithm's potential and increase its overall performance as in [22]. Further extensions such as long-short term memory or prioritized experience replay can be implemented and provide enhanced attributes for the proposed solution. Furthermore, future research should focus on an increased simulation to reality transfer. This can be accelerated significantly by multi-variable simulations which consider real-world uncertainties to minimize the existing implementation barriers. This also concerns the formulation of realistic objectives and reward functions, which do not only consider closed systems, but rather the interaction of the different actors.

To reduce general task complexities, further research can focus on a hierarchical RL frameworks, similar to the proposed rule selection framework in [39]. This circumvents the need for a single solution and distributes the varying objectives on distributed agents which are selected scenario-dependent according to pre-defined process criteria. Moreover, advanced edge functionalities can be implemented through cooperative learning and multi-agent architectures as discussed by [42] and [47]. Thus, the policy would not depend on the experience of a single agent, but would benefit from the totality of accumulated experience. The generation of a fleet intelligence would raise additional efficiencies and synergies in large and complex production systems and reduce the drawbacks of single-agent systems.

6. Conclusion

The purpose of this paper was to review existing applications of deep RL in production systems and to outline challenges and potential fields of future research. Based on a taxonomy framework, aggregated papers from three databases were narrowed to a final set of 38 papers and classified according to pre-defined criteria. It became apparent that deep RL has a broad application base in production scheduling, dispatching, and logistics, outperforming conventional algorithms in most cases and proving its ability to adapt to a wide variety of scenarios and handling production uncertainties. This not only optimized lead times, tardiness or WIP levels, but also reduced existing drawbacks of conventional methods such as high computation costs, limited adaptation capabilities, or high dependencies on human-based decisions. Nevertheless, only a few applications were assessed in reality, which makes further validation mandatory. More complex simulations that incorporate further uncertainties need to be conducted to reduce existing transfer barriers. Besides, additional consideration of optimization alternatives, such as more performant deep RL algorithms and extensions, should be considered to assess the full potential. Further research in collaborative and hierarchical multi-agent architectures and fleet intelligence approaches might also accelerate the deployment of deep RL and make it a reliable and robust optimization method for future distributed production systems.

Acknowledgements

The research presented in this paper has received funding by the Federal Ministry for Economic Affairs and Energy (BMWi), Germany (Az: 16KN086523:GeoFab – GIS2ALCM; Context-related utilization of GIS relations in life cycle management), via the VDI as project executing organization and the ZIM innovation program.

References

[1] ElMaraghy, H., AlGeddawy, T., Azab, A., and ElMaraghy, W., 2012. Change in Manufacturing – Research and Industrial Challenges, in: ElMaraghy, Enabling Manufacturing Competitiveness and Economic Sustainability, Springer Berlin Heidelberg, Berlin, Heidelberg, 2–9.

[2] Schmidtke, N., Behrendt, F., Thater, L., and Meixner, S., 2018. Technical Potentials and Challenges within Internal Logistics 4.0, in: 4th International Conference on Logistics Operations Management, IEEE, Le Havre, 1–10.

[3] Sutton, R. S., and Barto, A. G., 2017. Reinforcement Learning: An Introduction. The MIT Press, Cambridge, Massachusetts.

[4] Kumar, A., Dimitrakopoulos, R., and Maulen, M., 2020. Adaptive Self-Learning Mechanisms for Updating Short-Term Production Decisions in an Industrial Mining Complex. Journal of Intelligent Manufacturing 31 (7), 1795–1811.

[5] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D., 2018. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Nature 362 (6419), 1140-1144.

[6] Kang, Z., Catal, C., and Tekinerdogan, B., 2020. Machine Learning Applications in Production Lines: A Systematic Literature Review. Computers & Industrial Engineering 149, 106773.

[7] Liu, X., Xu, H., Liao, W., and Yu, W., 2019. Reinforcement Learning for Cyber-Physical Systems, in: 2019 IEEE International Conference on Industrial Internet, IEEE, Orlando, FL, USA, 318–327.

[8] Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S. F., Salwana, E., and Band, S. S., 2020. Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics. Mathematics 8 (10), 1640.

[9] Bellman, R., 1957. A Markovian Decision Process. Journal of Mathematics and Mechanics 6 (5), 679–684.

[10] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M., 2013. Playing Atari with Deep Reinforcement Learning, arXiv e-prints, arXiv:1312.5602.

[11] OpenAI, 2018. Welcome to Spinning Up in Deep RL! [Online]. Available: https://spinningup.openai.com. Accessed on: March 21 2021

[12] Rossit, D. A., Tohmé, F., and Frutos, M., 2019. Industry 4.0: Smart Scheduling. International Journal of Production Research 57 (12), 3802–3813.

[13] Tranfield, D., Denyer, D., and Smart, P., 2003. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. British Journal of Management 14 (3), 207–222.

[14] Thomé, A. M. T., Scavarda, L. F., and Scavarda, A. J., 2016. Conducting Systematic Literature Review in Operations Management. Production Planning & Control 27 (5), 408–420.

[15] Lohmer, J., and Lasch, R., 2020. Production Planning and Scheduling in Multi-Factory Production Networks: A Systematic Literature Review, International Journal of Production Research, 1–27.

[16] Lin, C., Deng, D., Chih, Y., and Chiu, H., 2019. Smart Manufacturing Scheduling With Edge Computing Using Multiclass Deep Q Network. IEEE Transactions on Industrial Informatics 15 (7), 4276–4284.

[17] Hubbs, C. D., Li, C., Sahinidis, N. V., Grossmann, I. E., and Wassick, J. M., 2020. A Deep Reinforcement Learning Approach for Chemical Production Scheduling. Computers & Chemical Engineering 141, 106982.

[18] Wu, C.-X., Liao, M.-H., Karatas, M., Chen, S.-Y., and Zheng, Y.-J., 2020. Real-Time Neural Network Scheduling of Emergency Medical Mask Production during COVID-19. Applied Soft Computing 97, 106790.

[19] Park, J., Chun, J., Kim, S. H., Kim, Y., and Park, J., 2021. Learning to Schedule Job-Shop Problems: Representation and Policy Learning Using Graph Neural Network and Reinforcement Learning. International Journal of Production Research, 1–18.

[20] Han, B.-A., and Yang, J.-J., 2020. Research on Adaptive Job Shop Scheduling Problems Based on Dueling Double DQN. IEEE Access 8, 186474–186495.

[21] Zhou, T., Tang, D., Zhu, H., and Wang, L., 2021. Reinforcement Learning With Composite Rewards for Production Scheduling in a Smart Factory. IEEE Access 9, 752–766.

[22] Baer, S., Turner, D., Mohanty, P., Samsonov, V., Bakakeu, R., and Meisen, T., 2020. Multi Agent Deep Q-Network Approach for Online Job Shop Scheduling in Flexible Manufacturing, in: 2020 International Conference on Manufacturing System and Multiple Machines, Tokyo, Japan, 1-9.

[23] Lee, S., Cho, Y., and Lee, Y. H., 2020. Injection Mold Production Sustainable Scheduling Using Deep Reinforcement Learning. Sustainability 12 (20), 8718.

[24] Rummukainen, H., and Nurminen, J. K., 2019. Practical Reinforcement Learning - Experiences in Lot Scheduling Application. IFAC-PapersOnLine 52 (13), 1415–1420.

[25] Palombarini, J. A., and Martinez, E. C., 2018. Automatic Generation of Rescheduling Knowledge in Socio-Technical Manufacturing Systems Using Deep Reinforcement Learning, in: 2018 IEEE Biennial Congress of Argentina, IEEE, San Miguel de Tucumán, Argentina, 1–8.

[26] Palombarini, J. A., and Martínez, E. C., 2019. Closed-Loop Rescheduling Using Deep Reinforcement Learning,IFAC-PapersOnLine, 52 (1), 231–236.

[27] Zhu, H., Li, M., Tang, Y., and Sun, Y., 2020. A Deep-Reinforcement-Learning-Based Optimization Approach for Real-Time Scheduling in Cloud Manufacturing, IEEE Access, 8, 9987–9997.

[28] Zhou, L., Zhang, L., and Horn, B. K. P., 2020. Deep Reinforcement Learning-Based Dynamic Scheduling in Smart Manufacturing,Procedia CIRP, 93, 383–388.

[29] Hu, L., Liu, Z., Hu, W., Wang, Y., Tan, J., and Wu, F., 2020. Petri-Net-Based Dynamic Scheduling of Flexible Manufacturing System via Deep Reinforcement Learning with Graph Convolutional Network, Journal of Manufacturing Systems, 55, 1–14.

[30] Liu, C., Chang, C., and Tseng, C., 2020. Actor-Critic Deep Reinforcement Learning for Solving Job Shop Scheduling Problems, IEEE Access, 8, 71752–71762.

[31] Baer, S., Bakakeu, J., Meyes, R., and Meisen, T., 2019. Multi-Agent Reinforcement Learning for Job Shop Scheduling in Flexible Manufacturing Systems, in: 2019 Second International Conference on Artificial Intelligence for Industries, IEEE, Laguna Hills, CA, USA, 22–25.

[32] Park, I., Huh, J., Kim, J., and Park, J., 2020. A Reinforcement Learning Approach to Robust Scheduling of Semiconductor Manufacturing Facilities. IEEE Transactions on Automation Science and Engineering 17 (3), 1420–1431.

[33] Chen, B., Wan, J., Lan, Y., Imran, M., Li, D., and Guizani, N., 2019. Improving Cognitive Ability of Edge Intelligent IIoT through Machine Learning. IEEE Network 33 (5), 61–67.

[34] Leng, J., Jin, C., Vogl, A., and Liu, H., 2020. Deep Reinforcement Learning for a Color-Batching Resequencing Problem. Journal of Manufacturing Systems 56, 175–187.

[35] Shi, D., Fan, W., Xiao, Y., Lin, T., and Xing, C., 2020. Intelligent Scheduling of Discrete Automated Production Line via Deep Reinforcement Learning. International Journal of Production Research 58 (11), 3362–3380.

[36] Xie, S., Zhang, T., and Rose, O., 2019. Online Single Machine Scheduling Based on Simulation and Reinforcement Learning, 18. ASIM Fachtagung Simulation in Produktion und Logistik, Chemnitz, 59-68.

[37] Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., and Kyek, A., 2018. Optimization of Global Production Scheduling with Deep Reinforcement Learning. Procedia CIRP 72, 1264–1269.

[38] Dittrich, M.-A., and Fohlmeister, S., 2020. Cooperative Multi-Agent System for Production Control Using Reinforcement Learning. CIRP Annals 69 (1), 389–392.

[39] Heger, J., and Voß, T., 2020. Dynamically Changing Sequencing Rules Wirth Reinforcement Learning in a Job Shop System with Stochastic Influences. Proceedings of the 2020 Winter Simulation Conference, 1608–1618.

[40] Luo, S., 2020. Dynamic Scheduling for Flexible Job Shop with New Job Insertions by Deep Reinforcement Learning. Applied Soft Computing 91, 106208.

[41] Altenmüller, T., Stüker, T., Waschneck, B., Kuhnle, A., and Lanza, G., 2020. Reinforcement Learning for an Intelligent and Autonomous Production Control of Complex Job-Shops under Time Constraints. Production Engineering 14 (3), 319–328.

[42] Stricker, N., Kuhnle, A., Sturm, R., and Friess, S., 2018. Reinforcement Learning for Adaptive Order Dispatching in the Semiconductor Industry. CIRP Annals 67 (1), 511–514.

[43] Kuhnle, A., Kaiser, J.-P., Theiß, F., Stricker, N., and Lanza, G., 2020. Designing an Adaptive Production Control System Using Reinforcement Learning. Journal of Intelligent Manufacturing 32, 855–876.

[44] Silva, T., and Azevedo, A., 2019. Production Flow Control through the Use of Reinforcement Learning. Procedia Manufacturing 38, 194–202.

[45] Waschneck, B., Reichstaller, A., Belzner, L., Altenmuller, T., Bauernhansl, T., Knapp, A., and Kyek, A., 2018. Deep Reinforcement Learning for Semiconductor Production Scheduling, in: 2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference, IEEE, Saratoga Springs, NY, USA, 301–306.

[46] Zheng, S., Gupta, C., and Serita, S., 2019. Manufacturing Dispatching Using Reinforcement and Transfer Learning, in: 2019 Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 655–671.

[47] Malus, A., Kozjek, D., and Vrabič, R., 2020. Real-Time Order Dispatching for a Fleet of Autonomous Mobile Robots Using Multi-Agent Reinforcement Learning. CIRP Annals 69 (1), 397–400.

[48] Feldkamp, N., Bergmann, S., and Strassburger, S., 2020. Simulation-Based Deep Reinforcement Learning for Modular Production Systems. Proceedings of the 2020 Winter Simulation Conference, 1596–1607.

[49] Hu, H., Jia, X., He, Q., Fu, S., and Liu, K., 2020. Deep Reinforcement Learning Based AGVs Real-Time Scheduling with Mixed Rule for Flexible Shop Floor in Industry 4.0. Computers & Industrial Engineering, 149, 106749.

[50] Kim, J.-B., Choi, H.-B., Hwang, G.-Y., Kim, K., Hong, Y.-G., and Han, Y.-H., 2020. Sortation Control Using Multi-Agent Deep Reinforcement Learning in N-Grid Sortation System, Sensors, 20 (12), 3401.

[51] Hildebrand, M., Andersen, R. S., and Bøgh, S., 2020. Deep Reinforcement Learning for Robot Batching Optimization and Flow Control. Procedia Manufacturing 51, 1462–1468.

[52] Xia, K., Sacco, C., Kirkpatrick, M., Saidy, C., Nguyen, L., Kircaliali, A., and Harik, R., 2020. A Digital Twin to Train Deep Reinforcement Learning Agent for Smart Manufacturing Plants: Environment, Interfaces and Intelligence. Journal of Manufacturing Systems 58, 210–230.

[53] Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., and De Freitas, N., 2016. Dueling Network Architectures for Deep Reinforcement Learning. Proceedings of the 33rd International Conference on Machine Learning, 1995–2003.

Biography

Marcel Panzer, M.Sc. (*1994) studied mechanical engineering at the Karlsruhe Institute of Technology and has been working as a research assistant at the University of Potsdam, Chair of Business Informatics, esp. Process and Systems since 2020. His research is focused on production planning and control.

Benedict Bender, Dr. (*1989) studied business informatics at the University of Potsdam, the Humboldt University of Berlin as well as the University of St. Gallen. His research interests include Industry 4.0 and aspects of IT security and privacy. Furthermore, he deals with digital platforms and business ecosystems.

Norbert Gronau, Univ.-Prof. Dr.-Ing. (*1964) studied mechanical engineering and business administration at the Technical University of Berlin (TU). In 1994, he received his doctorate in the Department of Computer Science (TU). Up to March 2000, he was head of the teaching and research group Production-Oriented Business Information Systems at the TU Berlin. He was head of the Department of Business Information Systems at the University of Oldenburg from 2000 to 2004. Since 2004 he holds the chair of Business Informatics, Processes and Systems.