Application of Reinforcement Learning for the Design and Optimization of Pass Schedules in Hot Rolling

C. Idzik(✉), J. Gerlach, J. Lohmar, D. Bailly, and G. Hirt

Institute of Metal Forming, RWTH Aachen University, Intzestr. 10, 52072 Aachen, Germany
christian.idzik@ibf.rwth-aachen.de

Abstract. About 95% of all steel products are rolled at least once during their production. Thus, any further improvement of the already highly optimized rolling process, for example reduction of energy consumption, has a significant impact. Currently, most rolling processes are designed by experts based on their knowledge and heuristics using fast analytical rolling models (FRM). However, due to the complex interactions between the processing constraints e.g. machine limits, the process parameters as well as the product properties, these manual process designs often focus on a single optimization objective. Here, novel methods such as reinforcement learning (RL) can detect complex correlations between chosen parameters and achieved objectives by interacting with an environment i.e. FRM. Therefore, this contribution demonstrates the potential of coupling RL and a FRM for the design and multiple objective optimization of rolling processes. Using FRM data e.g. the microstructure evolution, the coupled approach learns to map the current state, such as the height, to process parameters in order to maximize a numerical value and thereby optimize the process. For this, an objective function is presented that satisfies all (technical) constraints, leads to desired material properties including microstructural aspects and reduces the energy consumption. Here, two RL algorithms, DQN and DDPG, are used to design and optimize pass schedules for two use cases (different starting and final heights). The resulting pass schedules achieve the desired goals, for example, the desired grain size is achieved within 4 µm on average. These meaningful solutions can prospectively enable further improvements.

Keywords: Hot Rolling · Process Design · Reinforcement Learning

1 Introduction

In recent years, increased efforts have been made to further optimize energy-intensive processes such as hot rolling and thus make them more sustainable [1]. Hot rolling is one of the most common forming processes for the production of flat metal products with optimized dimensions and properties. About 95% of steel products are rolled at least once during their production as noted by Allwood et al. [1].

J. Lohmar: Deceased

M. Liewald et al. (Eds.): WGP 2022, LNPE, pp. 71–80, 2023.
https://doi.org/10.1007/978-3-031-18318-8_8

Consequently, even small optimizations have a major impact on global energy and material consumption. One major influencing factor regarding the process efficiency in the (hot) rolling process is the process design. Hot rolling processes generally consist of several rolling steps, called passes. During each pass the material is moved through at least one pair of rotating rolls and meanwhile deformed. The pass schedule defines all the parameters for the whole process, e.g. the height-reduction of each pass.

The pass schedule has to guarantee the defined geometric dimensions such as the final height and material properties such as a specific microstructure (grain size). Moreover, it has to consider rolling mill limitations, e.g. maximum allowable rolling force and torque, as well as economic aspects e.g. rolling time and ecological aspects e.g. the energy consumption. Typically, experts design pass schedules based on their knowledge, iterative heuristics and support of predictive models or simulations. However, these approaches often do not optimize for multiple objectives.

Here objective approaches that enable multi-objective optimization can support the experts to find optimized pass schedules. One possible solution is to combine methods of Machine Learning (ML) with physical process models. Scheiderer et al. [2] showed that ML methods, in form of Reinforcement Learning (RL), can design pass schedules while accounting for multiple objectives.

Therefore, in this paper RL algorithms are coupled with a fast rolling model (FRM) to design and optimize pass schedules for the hot rolling process. First, an overview of fast rolling models, approaches for pass schedule design and ML respectively RL is given in the state of the art. In this context, some application of RL for process optimization is shown. Next, the coupling between the RL algorithms and the FRM is detailed. In this paper, the reinforcement learning algorithms Deep Q-Network (DQN) [3] and Deep Deterministic Policy Gradient (DDPG) [4], are used. Additionally, the reward function for adhering to the specified tolerances, machine limits and total energy consumption to enable objective evaluations of pass schedules is presented. Subsequent, the results obtained for two trainings are presented and discussed. Finally, the results are summarized and an outlook regarding the next steps is given.

2 State of the art

This chapter presents the background information and state of the art of the used methods and approaches starting with fast rolling models and an overview of different design approaches of hot rolling processes. Next, ML and more specifically the RL is briefly introduced including current research on using RL for process optimization.

2.1 Fast Rolling Models

Here, conventional FE simulations, which take minutes or hours to run, are not suited because the total computation time of the optimization task depends on the underlying model that should compute as fast as possible. Hence, fast rolling models are used that usually consist of semi-empirical equations derived via mechanical simplifications or physical principles. Several fast rolling models have been developed in the past. Beynon and Sellars [5] present a rolling model called SLIMMER that is able to describe the microstructure evolution and predicts roll force and torque during multi-pass hot rolling. Inspired by their work, Seuren et al. [6] and Lohmar et al. [7] present a model that allows the prediction of force, torque, recrystallized fraction and grain size, among others, including a height resolution and the influence of shear. Another rolling model was developed by Jonsson [8]. It calculates, inter alia, strain, precipitates and dynamic recrystallization in order to predict the ferrite grain size after hot rolling.

2.2 Pass Schedule Design and Optimization

There are several different approaches to design a pass schedule in research and industry [9]. Many researchers and companies developed and proposed approaches dealing with these challenges. For instance, Svietlichnyi and Pietrzyk [10] suggest to set the height reduction in each pass to the maximum value while Schmidtchen and Kawalla [11] propose to set aim for an even distribution of the rolling force. All these approaches are usually used to design a first version of a pass schedule, which then is further optimized with regards to not more than two objectives simultaneously. Typical optimization objectives are grain refinement [5], grain size uniformity [12] as well as maximizing yield strength (YS) and ultimate tensile strength (UTS) [13]. A comprehensive literature review on recent design and optimization approaches for rolling is described by Özgür et al. [14].

2.3 Machine Learning and Its Applications for Process Optimization

Machine learning (ML) describes the ability of algorithms to detect patterns (in data) and learn from their inferences. Rosenblatt [15] proposed a probabilistic model that aims to replicate the functions of biological neurons. These artificial neurons transform an input vector into a scalar output by calculating a weighted sum out of all the input elements and then feeding them through an activation or transfer function. By systematically adjusting the weights, desired results are reproducible. Thus, multiple layers of these artificial neurons, called Artificial Neural Networks (ANN), can learn non-linear interrelations. These ANNs are used in different ML categories like supervised, unsupervised and reinforcement learning. Supervised learning needs a training set of labeled data and is used for classification and regression [16]. The goal of unsupervised learning is to find hidden structures in unlabeled data by clustering and dimension reduction.

The third category, reinforcement learning (RL), is an interacting approach where the algorithm learns by mapping state to actions to maximize a numerical reward [16]. One of the first ideas to use RL in manufacturing was published in 1998 [17]. However, according to Wuest et al. [18], RL is not widely applied in manufacturing, yet.

For hot rolling, Scheiderer et al. [2] published a RL approach that can design pass schedules while accounting for multiple objectives. The authors used a database with simulation data to design and optimize pass schedules. Moreover, Gamal et al. [19] demonstrated that RL in combination with process data identifies model parameters and thus improve predictions for bar and wire hot rolling processes.

2.4 Methodology: Coupling Reinforcement Learning and Fast Rolling Model

In this chapter, the coupling between a FRM and two RL algorithms is shown. For this purpose, the necessary components are described individually. First the used FRM is described, after which the RL algorithms and the optimization objectives are presented.

2.5 Fast Rolling Model

For the coupling with RL an already existing FRM is taken. The FRM, developed at the Institute of Metal Forming (IBF), is described in detail by Seuren et al. [6] and Lohmar et al. [7]. It is based on the slab method and consists of several modules, allowing the prediction of the deformation, the temperature and microstructure evolution as well as rolling forces and torques. In order to describe the material behavior, semi-empirical equations are used. For instance, the microstructure evolution is modelled using static recrystallization and grain growth equations from Beynon and Sellars [5].

The material used in this investigation is S355 since the model parameters are calibrated for this type of steel. S355 is a structure steel according to the European structural steel standard EN 10025–2:2004. The chemical composition is given in Table 1 and was determined using an optical emission spectrometer. The thermal boundary conditions used in the model are in accordance with values typically found in technical literature or were determined experimentally earlier**.**

Table 1. Chemical composition of S355

	C	Si	Mn	P	S	Cu	Al
Weight in %	0,1	0,3	1,6	0,012	0,001	0,2	0,02

2.6 Reinforcement Learning Algorithms

RL is a trial-and-error interaction approach. It consists of an agent, an environment, which represents the problem, and a set of actions. For each discrete time step, the agent perceives the environmental state and carries out actions, resulting in changes of the environment. Based on the resulting state a numerical value called reward is calculated. The reward assesses the goodness of the performed actions. The value can either be positive or negative, representing a reward or a punishment, respectively. The goal of RL is to choose the actions such that a maximum reward is obtained.

There are numerous RL algorithms available and described in the literature [16]. In this paper, the focus is set on two well-known and established RL algorithms used for process optimizations, Deep Q-Networks (DQN) [3] and Deep Deterministic Policy Gradient (DDPG) [4]. Both algorithms use ANNs to learn the relationships between actions, new states and the resulting rewards. Through interaction, it generates the data samples for training itself. But the samples are correlated, they are stored in a so-called experience buffer from which random samples, called a mini-batch, are taken.

DQN uses one ANN as an approximator for the state-value function which indicates how beneficial it is for an agent to perform a particular action in a state. This state-value function, often referred to as Q function, describes the cumulative long-term reward.

DDPG is very similar to DQN, except for the fact that is uses two ANNs, the critic and the actor network. The critic network has the same task as the DQN, but instead of using just the Q-value to improve action selection, DDPG uses the Q-values to learn a policy using the actor network. This allows to solve more complex problems with even a continuous action space. The general structure of the coupling between the RL and the FRM is exemplarily shown for DDPG in Fig. 1.

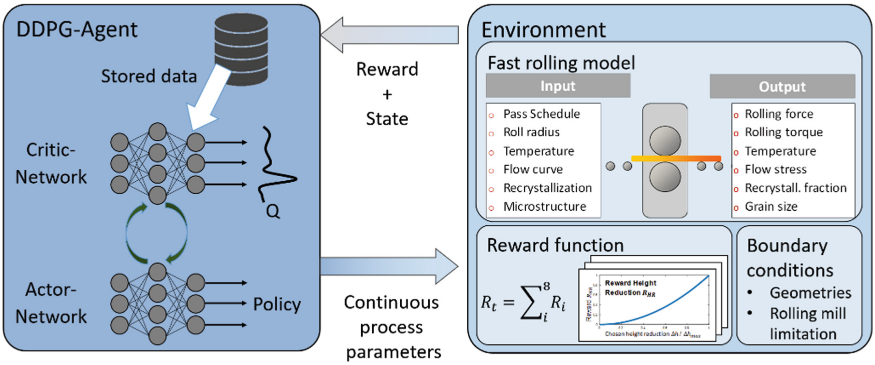


Fig. 1. Coupling of the DDPG algorithm with the FRM

In order to establish a better comparability between the two RL algorithms, the same hyperparameters were used as far as possible. In both DQN and DDPG, the ANNs have a similar structure. The critic networks in DQN and DDPG have four hidden layers with 200 neurons each while the actor network of the DDPG algorithm contains two hidden layers with 200 neurons each. Table 2 shows the most relevant hyperparameters of the two algorithms. They correspond to typical values from the literature and were chosen equally for both algorithms as far as possible.

Table 2. Hyperparameters for the used DQN and DDPG

Hyper-parameter	Discount factor	Learning rate	Experience buffer length	Mini-batch size
	0.95	1e−3	1e5	64

2.7 Optimization Objectives

As mentioned before, the goal is to optimize hot rolling with respect to multiple objectives. For this purpose, the optimization objectives are first defined and converted into a corresponding reward function R. The objectives should be based on real conditions in production and take into account factors such as product properties (height, grain size), limitations (rolling mill limits, final temperature) and process efficiency (energy consumption, process time). Table 3 shows the objectives considered here.

Table 3. Optimization objectives considered in the reward function

Reward component	Objective of optimization
Height R_H	Reward if h_{target} is reached, punish if $h_c < h_{target}$
Grainsize R_d	Reward when the target range of d is reached
Force, torque R_F, R_M	Exploit the rolling mill limit but not exceed it
Temperature R_ϑ	Maximizing final temperature
Energy consumption R_E	Minimizing total energy consumption
Height reduction R_{HR}	Minimizing the pass number
Interpass time R_I	Minimizing the process time

After defining the optimization objectives and thus individual reward components, the actual reward function R is defined, see Eq. 1. An intuitive possibility is to set up a weighted sum of the reward components. This allows easy prioritization so that in the context of this publication, product properties such as the final height R_H and microstructure in terms of the grain size R_d are prioritized and accordingly weighted stronger. The other components are weighted equally, as there is no further preference here. Weighting is necessary otherwise the desired properties like the final height and grain size are achieved within the tolerance. A simple sum of all reward components resulted in pass schedules that did not simultaneously achieve the target height and the target grain size. Through several trials, the weights for R_H and R_d were adjusted so that first the target height and then target grain size was reached.

$$R_0 = 5 * R_H + 3 * R_d + R_E + R_{HR} + R_M + R_F + R_I + R_\vartheta \tag{1}$$

In general, the reward components can be divided into three purposes. The first purpose is to reach a certain target value or area. This applies to the reward components of height R_H, grain size R_d and rolling force R_F as well as torque R_M. In Fig. 2, the definition of the R_H is presented exemplarily. R_d, R_F and R_M are defined similarly.

In addition to achieving certain targets within defined tolerances, there are also objectives that need to be minimized (R_E, R_I) or maximized (R_{HR}, R_ϑ). As in the previous example, the reward components are defined in an uniform way as shown in Fig. 3.

3 Results and Discussion: Designed Pass Schedules

In this chapter, the final pass schedules for two use cases are described, discussed especially in the context of possible novel approaches to design pass schedules. The use cases

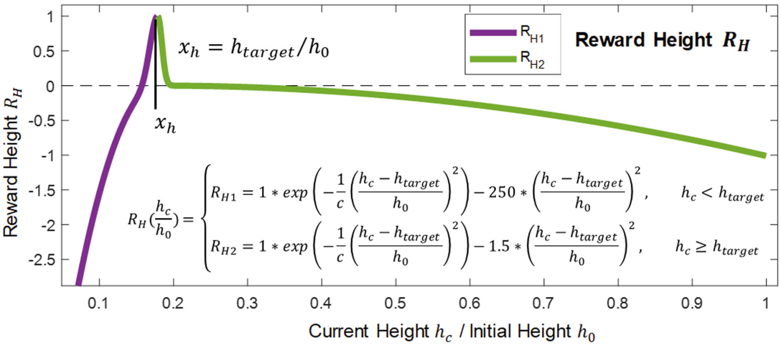


Fig. 2. Reward component for the height

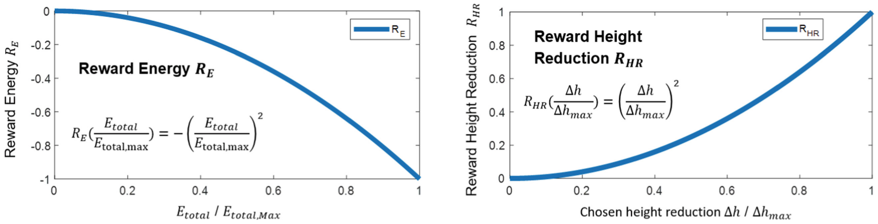


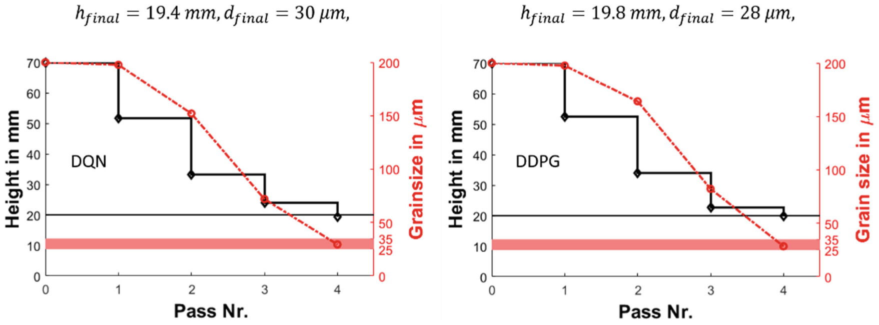
Fig. 3. Reward components for the energy consumption (left) and height reduction (right)

are selected allowing experimental rolling on the laboratory rolling mill at IBF. The initial and final parameters of the two cases can be found in Table 4. Both were trained in 20,000 iterations with the DQN and the DDPG algorithm. These trainings took about 24 h on an average computer (CPU: Intel Xeon E3-1270). If 20,000 FE simulations were calculated instead of the FRM, the training would much longer. The agents could choose the height reduction [3–18.45 mm] and inter-pass time [5–15 s]. The remaining process parameters were held constant (for example the rolling velocity is fixed at 250 mm/s) but can generally be included in the optimization, as well.

Table 4. Initial and target parameters for the two use cases

Use case	Initial parameters				Targets	
	h_0 in mm	b_0 in mm	l_0 in mm	ϑ_0 in °C	h_{target} in mm	d_{target} in µm
1	70	160	320	1100	20 ± 0.4	30 ± 5
2	140	180	500	1200	25 ± 0.5	30 ± 5

First, the resulting pass schedules of the first use case are presented and discussed, see Fig. 4. The pass schedule laid out by DQN consists of a total of four passes, which is the minimum number of passes possible. The most important targets regarding the final properties were achieved only to a limited extent. The final height h_{final} 19.4 mm is slightly outside the desired tolerance, whereas the average grain size d_{final} 30 μm was perfectly achieved. The pass schedule laid out by DDPG consists of a total of four passes and results in targets which are within the desired tolerances (h_{final} 19.8 mm, d_{final} 28 μm). Otherwise, both pass schedules are very similar. In both cases, pass schedules do not show a novel rolling strategy, e.g. the height reduction is reduced with each pass. This practice is very similar to that of experts. Experts would also start with as few passes as possible, starting with large height reductions at the beginning and decreasing the height reduction with each pass. Therefore, the two automatically designed pass schedules do not differ significantly from those designed by experts.


$h_{final} = 19.4\,mm, d_{final} = 30\,\mu m,$

$h_{final} = 19.8\,mm, d_{final} = 28\,\mu m,$

Fig. 4. Final pass schedules for the first use case trained by DQN (left) and DDPG (right)

Comparing the obtained pass schedules for the second use case, see Fig. 5, a lot of similarities can be observed here as well. In both cases, the pass schedules consist of eight passes (minimum number of passes possible), the target value for h_{final} (24.7 and 24.9 mm) is reached while d_{final} (26 and 23 μm) is not, although the results lie very close to the tolerance area. In contrast to the first case, these two pass schedules with the very small height reduction in the last pass show a conspicuousness compared to typical pass schedule designed by experts. Such small height reductions in the last pass would very probably not be suggested by an expert, as it has no added value in process terms. Experts would reduce the height reduction as described above with each pass, but they would distribute it more evenly and not just reduce it by a few mm.

All obtained results show that the concept of coupling a FRM with RL is very promising. Rollable pass schedules were designed and the desired objectives were (almost) reached. However, it is evident that despite stronger weighting in the reward, the final heights are not hit perfectly. Currently, the final height needs to be adjusted slightly to achieve the targets perfectly. Interestingly, there are no noticeable differences between the two RL-algorithms (DQN, DDPG) in the final designed pass schedules for both use cases. Both algorithms can be used for pass plan design.

4 Conclusion

In this paper, two hot rolling processes were designed automatically to demonstrate the promising potential of coupling a process model, FRM and RL. Two different RL

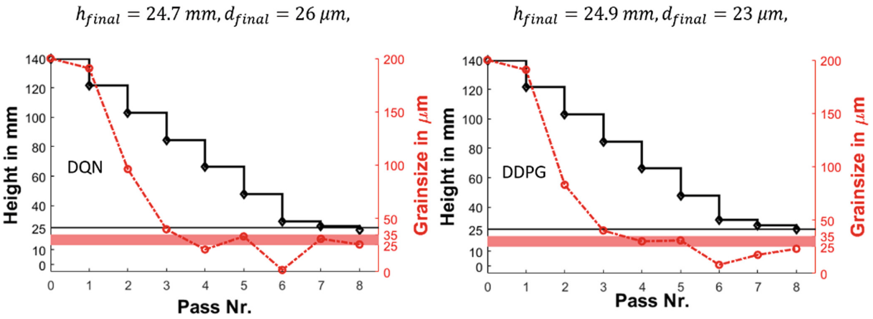


Fig. 5. Final pass schedules for the second use case trained by DQN (left) and DDPG (right)

algorithms, DQN and DDPG, were used for the training to test and compare them. The designed pass schedules after training show that, regardless of the use case and specific RL algorithm used, the coupled approach leads to rollable pass schedules that successfully achieve goals for the most part. In both presented use cases DQN and DDPG lay out similar final pass schedules. It is noticeable that for the first use case ($h_{final} = 20$ mm) presented, both algorithms achieve very good pass schedules in terms of final height and grain size. The average deviation in the final height Δh is 0.4 mm and in the mean grain size Δd is just 1 µm. In the second use case, requiring twice the number of passes compared to the first one, Δh was better (0.2 mm), but Δd was worse (5.5 µm). Comparing the designed pass schedules with typical pass schedules laid out by experts, no novel rolling strategy can be identified. The pass schedules designed by the RL algorithms apply high height reductions at the beginning and reduce them with each pass. In the future, the approach will be extended for online process adaption.

Acknowledgement. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC-2023 Internet of Production – 390621612. We thank Kuan Wang for supporting us in conducting the trainings.

We heartily thank our colleague Dr.-Ing. Johannes Lohmar for his encouragement and for his timely support, guidance and suggestions during this project work.

References

1. Allwood, J.M., Cullen, J.M., Carruth, M.A.: Sustainable materials. With both eyes open; [future buildings, vehicles, products and equipment - made efficiently and made with less new material]. UIT Cambridge, Cambridge (2012)
2. Scheiderer, C., et al.: Simulation-as-a-service for reinforcement learning applications by example of heavy plate rolling processes. Proc. Manuf. **51**, 897–903 (2020). https://doi.org/10.1016/j.promfg.2020.10.126
3. van Hasselt, H., Guez, A., Silver, D.: Deep reinforcement learning with double Q-learning (2015)
4. Silver, D., Lever, G., Heess, N., Degris, T,. Wierstra, D., Riedmiller, M.: Deterministic Policy Gradient Algorithms Proceedings of the 31 st International Conference on Machine Learning, 32. Aufl, Beijing, China, S 387–395 (2014)

5. Beynon, J.H., Sellars, C.M.: Modelling Microstructure and Its Effects during Multipass Hot Rolling. Iron Steel Inst. Jap. **32**(3), 359–367 (1992)
6. Seuren, S., Bambach, M., Hirt, G., Heeg, R., Philipp, M.: Geometric factors for fast calculation of roll force in plate rolling. In: Zhongguo-Jinshu-Xuehui (Hrsg) 10th International Conference on Steel. Metallurgical Industry Press, Beijing (2010)
7. Lohmar, J., Seuren, S., Bambach, M., Hirt, G.: Design and application of an advanced fast rolling model with through thickness resolution for heavy plate rolling. In: Guzzoni, J., Manning, M. (Hrsg) 2nd International Conference on Ingot Casting Rolling Forging. ICRF (2014)
8. Jonsson, M.: An investigation of different strategies for thermo-mechanical rolling of structural steel heavy plates. ISIJ Int. **46**(8), 1192–1199 (2006). https://doi.org/10.2355/isijinternational.46.1192
9. Pandey, V., Rao, P.S., Singh, S., Pandey, M.: A calculation procedure and optimization for pass scheduling in rolling process. A Rew. 126–130 (2020)
10. Svietlichnyj, D.S., Pietrzyk, M.: On-line model for control of hot plate rolling. In: Beynon, J.H. (Hrsg) 3rd International Conference on Modelling of Metal Rolling Processes. IOM Communications, London, S 62–71 (1999)
11. Schmidtchen, M., Kawalla, R.: Fast Numerical simulation of symmetric flat rolling processes for inhomogeneous materials using a layer model—part I. Basic Theory. Steel Res. Int. **87**(8), 1065–1081 (2016). https://doi.org/10.1002/srin.201600047
12. Hong, C., Park, J.: Design of pass schedule for austenite grain refinement in plate rolling of a plain carbon steel. J. Mater. Process. Technol. **143–144**, 758–763 (2003). https://doi.org/10.1016/S0924-0136(03)00363-7
13. Chakraborti, N., Siva Kumar, B., Satish Babu, V., Moitra, S., Mukhopadhyay, A.: A new multi-objective genetic algorithm applied to hot-rolling process. Appl. Math. Model. **32**(9), 1781–1789 (2008). https://doi.org/10.1016/j.apm.2007.06.011
14. Özgür, A., Uygun, Y., Hütt, M.-T.: A review of planning and scheduling methods for hot rolling mills in steel production. Comput. Ind. Eng. **151**(20), 106606 (2021). https://doi.org/10.1016/j.cie.2020.106606
15. Rosenblatt, F.: The perceptron. A probabilistic model for information storage and organization in the brain. Psychol. Rev. **65**(6), 386–408. (1958). https://doi.org/10.1037/h0042519
16. Sutton, R.S., Barto, A.: Reinforcement Learning. An Introduction. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, London (2018)
17. Mahadevan, S., Theocharous, G.: Optimizing Production Manufacturing Using Reinforcement Learning FLAIRS conference, Bd **372**, S 377 (1998)
18. Wuest, T., Weimer, D., Irgens, C., Thoben, K.-D.: Machine learning in manufacturing. Adv. Chall. Appl. Prod. Manuf. Res. **4**(1), 23–45 (2016). https://doi.org/10.1080/21693277.2016.1192517
19. Gamal, O., Mohamed, M.I.P., Patel, C.G., Roth, H.: Data-driven model-free intelligent roll gap control of bar and wire hot rolling process using reinforcement learning. IJMERR 349–356 (2021). https://doi.org/10.18178/ijmerr.10.7.349-356