Data-Driven Model-Free Intelligent Roll Gap Control of Bar and Wire Hot Rolling Process Using Reinforcement Learning

Omar Gamal, Mohamed Imran Peer Mohamed, Chirag Ghanshyambhai Patel, and Hubert Roth
University of Siegen, Siegen, Germany
Email: omar.gamal@uni-siegen.de, {mohamed.pmohamed, Chirag Ghanshyambhai Patel}@ student.uni-siegen.de, hubert.roth@uni-siegen.de

***Abstract*— Modern bar and wire manufacturing plants are constantly seeking to achieve lower costs, higher product quality, higher efficiency, and greater flexibility, which in turn require a significant increase in the degree of automation. The lack of accurate models and measurements of process essential parameters affects the realization of innovative control strategies. Data-driven approaches have received a lot of attention from control engineering researchers owing to their outstanding performance as function approximators. Supervised, unsupervised, and reinforcement learning approaches have been successfully employed in system identification problems and control systems design. Reinforcement learning holds advantages over the other approaches owing to its ability to learn without having the desired ground truth state. In this paper, a data-driven model-free reinforcement learning algorithms are developed for model parameters identification and roll gab control of a bar and wire hot rolling process. The reinforcement learning algorithms are based on the Deep Deterministic Policy Gradients algorithm with the actor-critic structure. The validation results of the developed solutions showed high performance and the agents were able to generalize to unseen scenarios.**

***Index Terms*— hot rolling, bar and wire process, roll gap control, parameter estimation, reinforcement learning, Deep Deterministic Policy Gradients**

I. Introduction

Profile rolling is considered to be one of the oldest forming processes in metal forming where products such as bar, wire, beams, etc. are produced. In bar and wire hot rolling process, the material is processed at a temperature above its recrystallization temperature, which reduces the rolling forces and increases the material forming capacity [1]. Nevertheless, the forming process is rather complex owing to the nature of roll passes used (e.g., diamond, round, oval, square, etc.), the rolling technology employed (e.g., two, three, and four roll technologies), and the wide range of input materials which leads to cross-sectional variations and thus deteriorate end product tolerances [2]. This imposes a lot of challenges to construct mathematical models that describe the underlying dynamics accurately.

Manuscript received January 4, 2021; revised March 11, 2021.

The forming process can be described by calibration methods and stress distribution in the roll gap. The calibration methods are used to calculate the roll gab and material geometry. It can be classified into regular, simple irregular, and complicated irregular calibration [1]. In simple irregular calibration, the equivalent rectangle method can be used for simplification based on Lendl [3] or Hensel [4] for two roll technology and C. Overhagen, and P. J. Mauk [5] for three roll technology. The information obtained from the calibration method along with stress distribution in the roll gap are used as a basis for material spread, roll force, and torque calculations. For the calculation of the roll force and torque, several models are available [5-8]. Most of the developed models, however, are transferred from the theory of flat rolling, where the roll pass is converted into a rectangle of equal area for simplification. These models are uncertain owing to their inaccurate representation of process physics as well as unmodelled dynamics.

The lack of accurate models and measurements of process essential parameters affects the realization of innovative control strategies for bar and wire hot rolling processes. Thus, an experienced operator is often used to control the entire process and intervene from time to time to adjust the process parameters. This makes the performance of the process entirely dependent on the experience of the operator which is prone to many limitations such as understanding the underlying process physics and making real-time decisions. This all puts engineers and operators of this plant in front of many challenges to cope with the higher degree of complexities and uncertainties of such highly nonlinear and coupled dynamic systems.

Dynamic system identification is concerned with model structure and parameter identification problems. Based on prior information, models can be categorized into white-box, gray-box, and black-box models [9]. The identified models are used for system output prediction and control in a wide variety of engineering applications. The identification process of complex dynamic systems, however, is challenging owing to their nonlinear, dynamic characteristics, and the lack of physical insight in many plants. This poses a major challenge to classical identification methods to obtain optimal results.

doi: 10.18178/ijmerr.10.7.349-356

Recent advances in deep learning have drawn researchers' attention in control engineering field to its outstanding performance as function approximators. Supervised, and unsupervised methods have been employed in system identification problems and control systems design. In [10], the authors proposed an unsupervised learning approach for the identification of Piece Wise Smooth Hybrid Systems (PWS-HS). The authors in [11] designed three neural networks, each with three layers for system identification and optimal controller design of unknown discrete-time nonlinear systems. W. Yu and X. Li [12] examined the stability of dynamic neural networks in the identification of nonlinear systems using passivity theory. In [13], the authors designed a neural network with a single hidden layer for the identification of discrete-time non-linear systems. Reinforcement Learning (RL) algorithms have also been successfully employed in system identification and controller design. In contrast to supervised and unsupervised learning, reinforcement learning learns by interacting with the environment and taking actions that maximize a cumulative reward [14]. In [15], the authors proposed a Continuous Action Reinforcement Learning Automata (CARLA) for the identification of multiple-input multiple-output (MIMO) systems. The authors in [16] designed a low-level hover controller for a quadrotor using model-based reinforcement learning. Other implementations of reinforcement learning in the identification and control of dynamic systems can be found in [17-22].

Motivated by the aforementioned problems and work, in this paper, novel reinforcement learning algorithms are developed for model parameter identification and roll gab control of a bar and wire hot rolling process. The main advantage of reinforcement learning is its ability to learn without having a model of the environment and the desired model output. To the best of our knowledge, this is the first employment of reinforcement learning methods in the identification and control of bar and wire hot rolling processes. The performance of the trained RL-Agents is evaluated using the regret metric. The identified process model is evaluated in simulation by measuring the overall identification error. Further, the effectiveness of the roll gap controller is tested in simulation.

II. Methods

The nature of the forming process in bar and wire hot rolling processes poses a major challenge to construct models that represent process dynamics accurately. In [23], we presented a parametric dynamic model for the finishing mill stand block in a bar and wire hot rolling process. The mill stand block consists of six individually driven mill stands with 3-roll technology. The model structure is built using the physical insights of the process, i.e. white-box model. The model parameters, however, are not precisely known, and thus they must be estimated using observations. In this work, gray-box model is adopted where reinforcement learning algorithms are used to compensate for model inaccuracies using measured data from the real plant. The algorithm is used to adapt three essential process parameters; namely kappa, roll force, and torque. The dataset consists of multiple time series for 100CR6 material, each represents a process parameter measurement, e.g. input material cross-sectional area, roll gab, roll force, motor toque, motor angular velocity, etc.

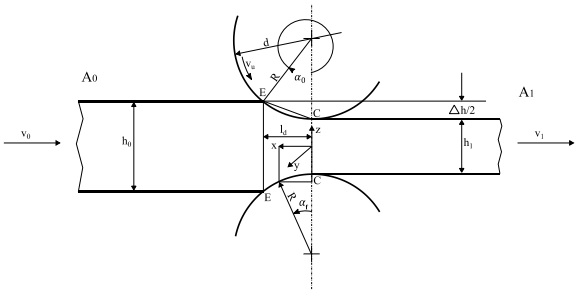

Figure 1. Roll gap geometry.

In bar and wire hot rolling process, there is a high demand for improved control concepts to reduce downtime and improve product quality and tolerances. The output material cross-sectional area and tolerances are highly influenced by many factors, e.g. roll gab of the mill stand, interstand tension, material temperature, etc. An experienced operator is often used in real plants to control the end product cross-sectional area and tolerances based on optimized setting values which are chosen based on experience. The optimized setting values include roll gab, motor angular velocity, and roll force values. In this paper, we present a data-driven model-free intelligent roll gab controller using reinforcement learning. The process optimized setting values are used as the initial setpoints for the process model.

To apply reinforcement learning methods to continuous systems, the algorithms must deal with continuous action and state spaces. Therefore, Deep Deterministic Policy Gradients (DDPG) algorithm with the actor-critic structure is employed, which can handle high-dimensional continuous action and state space. The performance of the trained RL-Agents is evaluated using the regret metric. The identified process model is evaluated in simulation by measuring the overall identification error. Further, the effectiveness of the roll gap controller is tested in simulation.

III. Bar and Wire Process Model

The process model presented in [23] is based on Lippmann and Mahrenholtz roll model [8], where the roll pass is converted into a rectangle of equal area for simplification. The roll gap and material geometry are calculated using the equivalent rectangle method [1]. The roll force P and torque M can be calculated, as in:

$$P=b\, k_{ep} \sqrt{R\, \Delta h} \cdot f_p. \quad (1)$$

$$M=b\, k_{eM}\, R\, \Delta h \cdot f_M. \quad (2)$$

Where b is the material width, R is the roll radius, Δh is the change in height, k_{ep} is the yield stress, k_{eM} is the mean yield stress for a forming step, and f_p, f_M are auxiliary functions, which are dependent on the

compressive stresses in the rolling direction, strain, and neutral point angle.

Throughout the rolling process, the rolled stock enters the mill stand with a cross-sectional area A_0 and exits with a cross-sectional area A_1. The reduction in cross-sectional area results from the applied pressure forces within the roll gap. The reduction in material cross-sectional area is followed by an increase in material speed. Throughout the forming zone, one encounters three different and important velocities: the circumferential speed of the rolls V_u, the input speed of the rolling stock V_0 and the exit speed V_1. In the entry section, the roll circumferential speed exceeds the input material speed V_0. The material speed then gradually increases as it progresses through the roll gap until it reaches the same speed as the rolls at the neutral point [1]. As the stock progresses beyond this point, its speed gradually increases beyond the roll speed. As there is no mass exchange with the environment, the mass remains constant before and after the forming process. For constant volume flow rate,

$$\dot{V} = V_0\, A_0 = V_1\, A_1 = \text{Constant}. \quad (3)$$

$$V_1 = (1 + \kappa)\, V_u. \quad (4)$$

$$\kappa = \frac{A_F}{A_1}\cos(\alpha_F) - 1. \quad (5)$$

The relationship between the circumferential speed of the rolls V_u and the exit speed of the rolling stock V_1 can be identified by (4) and (5), where A_F is the material cross-sectional area at the neutral point and α_F is the neutral point angle.

IV. Reinforcement Learning

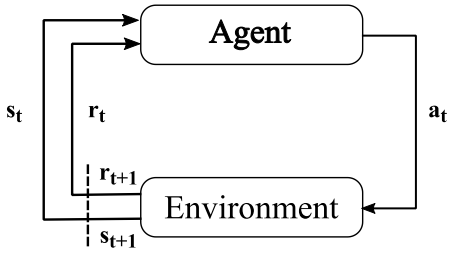

Figure 2. The learning scheme of the reinforcement learning algorithm.

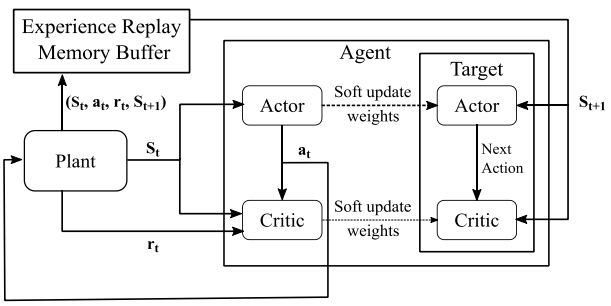

Figure 3. DDPG training architecture.

Reinforcement Learning (RL) is an area in machine learning in which the model (Agent) learns the optimal actions that maximize a cumulative reward R by interacting with the environment [24]. Fig. 2 illustrates the learning scheme of the reinforcement learning algorithm, where A_t is the action executed by the agent based on its current state S_t, and makes it transition to state S_{t+1}. The states can be discrete or continuous based on the system structure.

Reinforcement learning methods can be categorized into Model-Free and Model-Based algorithms. In Model-Based learning, the optimal policy is chosen based on the learned internal model of the system. On the other hand, Model-free learning uses experience for setting up the optimal policy without having an environment model. Based on the nature of the environment different RL algorithms can be chosen, e.g. Q-Learning, State-Action-Reward-State-Action (SARSA), Deep Q-Network (DQN), and Deep Deterministic Policy Gradient (DDPG). In contrast to DQN, Deep Deterministic Policy Gradient (DDPG) [25] can be employed in continuous action space. The algorithm uses actor-critic architecture where the actor outputs an action based on the input state. The critic function criticizes the action based on the state and reward. DDPG consists of four networks; namely actor, critic, target actor, and target critic networks, Fig. 3.

To evaluate the performance of the agents, regret metric is used. Regret is the difference between the learned policy cumulative reward and the cumulative reward for the optimal policy (6) [26]. In case of optimal policy, the maximum reward is considered for the regret calculation. The agent with a smaller regret value is the best.

$$R^T = \max_{a^T \in A} \frac{1}{T}\sum_{t=1}^{T} r(a^t) - \frac{1}{T}\sum_{t=1}^{T} r(\dot{a}^t). \quad (6)$$

Where R is the regret value calculated for an episode with T steps, $r(a^t)$ represents the reward gained for the optimal policy and $r(\dot{a}^t)$ represents the reward gained for learned policy.

V. Model Parameter Adaptation

Three essential parameters are adapted using DDPG algorithm; namely kappa, roll force, and torque. The DDPG Algorithm has an actor-critic network configuration. The actor-critic network of the kappa, roll force, and torque agents have the same architecture.

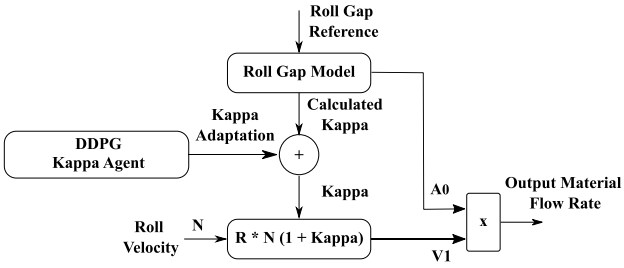

Figure 4. Kappa adaptation.

The actor-network takes an input of size 5X1 and consists of four fully connected layers with 80, 60, 40, and 1 unit respectively, each is followed by an activation layer. The first three activation layers have ReLU activation and the last layer has Tanh activation. The

critic network has two branches; namely State and Action branches. The state branch takes an input of size 5X1 and consists of two fully connected layers with 80, 60 units respectively and one ReLU activation layer. The action branch takes as an input the action taken by the actor-network with size 1X1 which is then followed by a fully connected layer with 60 units. The output of both branches, i.e. State and Action branches, are then concatenated and the output is given as an input to ReLU activation layer, which is followed then by a fully connected layer with one unit.

As the state and action spaces are continuous, a continuous reward function (7) is used, where T is the error tolerance and E is the error of the parameter being adapted, i.e. flow rate E_q, roll force E_F, and roll torque E_M [27].

$$R=\sqrt{\left(\frac{T}{50}\right)}-\sqrt{\left(\frac{|E|}{50}\right)}. \quad (7)$$

A. Kappa Adaptation

Since the material moves within the system as an elastic body and experiences spatial and temporal changes similar to flowing fluids, the conservation of mass law from fluid mechanics holds (3). In the process model presented earlier the material output volume flow rate deviates from the input material flow rate. To ensure a constant volume flow rate, Kappa is adapted with the help of DDPG algorithm. In this case, the kappa agent provides an action based on the observed states which is used to adapt the kappa value calculated by the model, see Fig. 4. The observed states are input material cross-sectional area, input material speed, motor speed, roll gap, and flow rate error. Five agents in total are used to adapt the kappa parameter of the first five mill stands in the mill stand block. The sixth mill stand is not used in our dataset.

The agents are configured to collect the maximum reward, i.e. long-term reward over an episode by setting discount factor to 0.99. The actor and critic networks are trained from the random samples stored in the experience replay buffer. The experience replay buffer size is set to 1000000. And the mini-batch size is set to 256. The target smooth factor (τ) to transfer the weights to the target network is 0.001. The exploration factor of the agent is determined using the noise factor which is set to 0.01. Each episode starts when the material enters the mill stand and terminates when the agent has generated actions for 1.5 secs (1500 steps). The training is stopped when the agent achieves an average of 50 episodes' reward of 100 or above.

To evaluate the performance of the trained Kappa agents, they are deployed in simulation, see Fig. 5. The identified process model is evaluated by measuring the overall identification error (8), where $\|xref\|_2 and \|x\|_2$ are the norms of the identified and true system [28]. The kappa agents achieved an identification error close to zero.

$$E_r = \frac{\|xref\|_2 - \|x\|_2}{\|x\|_2}. \quad (8)$$

To evaluate the model performance further, regret is calculated for the five agents. The five agents achieved regret score of 0.24, 0.87, 1.38, 0.55, and 0.7 respectively.

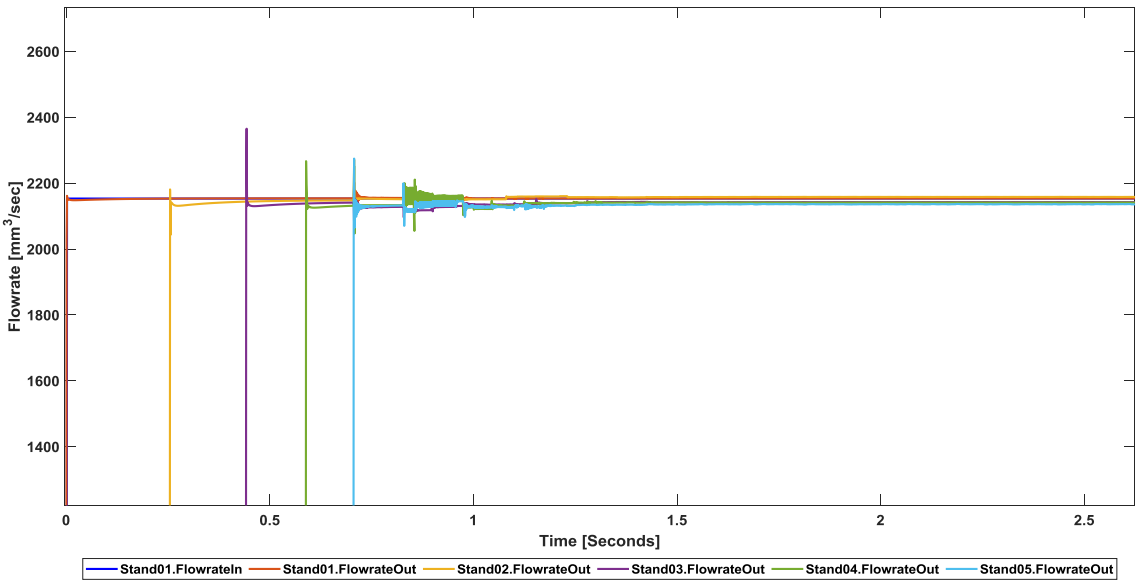

Figure 5. Adapted flow rate using DDPG Kappa agents.

B. Roll Force Adaptation

In bar and wire hot rolling process, the material is subjected to different forces within the roll gap; namely the resultant pressure exerted over the area of contact and resultant friction forces. The resultant force, i.e. roll force is normal to the area of contact. An accurate prediction of the roll force depends mainly on the stress-strain distribution within the roll gap.

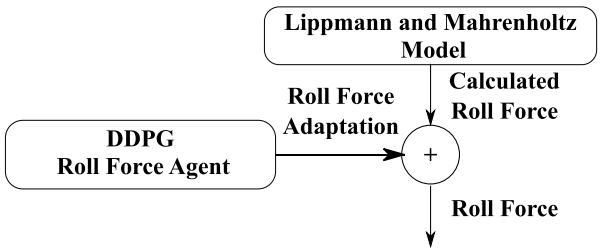

Figure 6. Roll Force adaptation.

To adapt the deviation observed in the calculated roll force in each mill stand, five DDPG agents are used. In this case, the roll force agent provides an action based on the observed states which are used to adapt the roll force value calculated by the model, see Fig. 6. The observed states are input material temperature, input material speed, actual motor speed, input material cross-sectional area, and the roll gap.

The roll force agent training configuration remains the same as kappa agents with a slight variation in the discount factor, which is set to 0.01 and exploration is set to 0.9 with a decay rate of 0.02. All the agents are trained for 1 sec (1000 steps). The agent training is stopped when the agent achieves the average of 5 episodes' reward of 200 or above. The agent training is started when the materials enter the respective mill stand.

To evaluate the performance of the trained Roll force agents, they are deployed in simulation. The agents showed high performance with a small deviation from the measured roll force, see Fig. 7. The roll force agents achieved an identification error close to zero. To evaluate the model performance further, regret is calculated for the five agents. The five Roll force agents achieved regret score of 1.32, 0.61, 0.97, 0.93, and 0.87 respectively.

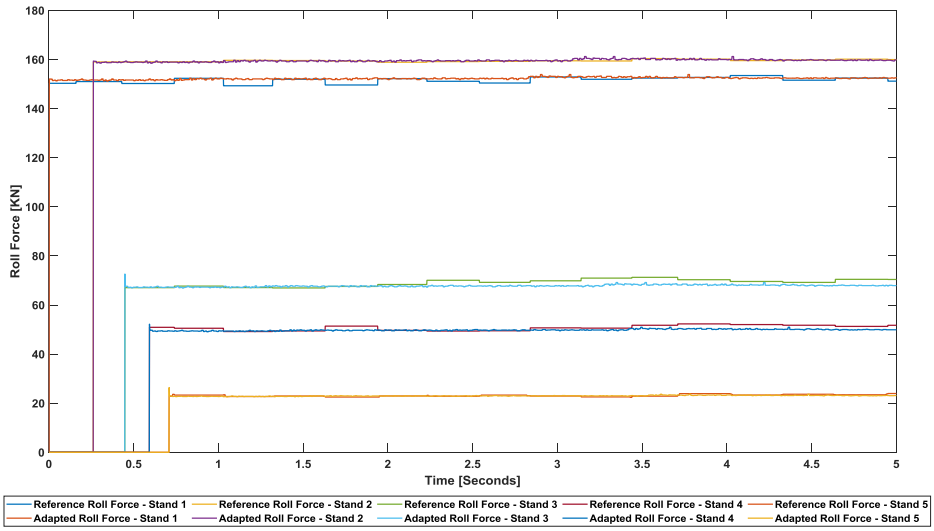

Figure 7. Adapted roll force using DDPG roll force agents.

C. Roll Torque Adaptation

The dataset obtained from the real plant does not have roll torque measurement data. To adapt the calculated roll torque (2) we used roll force measurement value to back-calculate the roll torque (9), where l_d is the compressed length and F_a is measured roll force.

$$M_r = 0.5 * l_d * F_a. \quad (9)$$

To adapt the deviation observed in the calculated roll torque in each mill stand, five DDPG agents are used. In this case, the roll torque agent provides an action based on the observed states, which are used to adapt the roll torque value calculated by the model, see Fig. 8. The observed states are input material temperature, input material speed, actual motor speed, input material cross-sectional area, and the roll gap. The roll torque agents' configuration remains the same as the roll force agents. The training is stopped when the agent achieves the average of 5 episodes' reward of 150 or above.

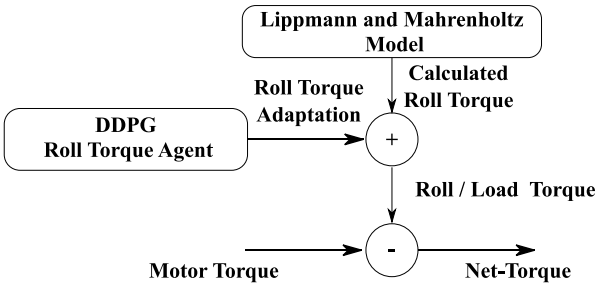

Figure 8. Roll torque adaptation.

To evaluate the performance of the trained roll torque agents, they are deployed in simulation. The agents showed high performance with a small deviation from the back-calculated roll torque, see Fig. 9. The roll torque agents achieved an identification error close to zero. To evaluate the model's performance further regret is calculated for the five agents. The five Roll torque agents achieved a regret score of 0.8, 0.9, 1.2, 1.0, and 0 respectively.

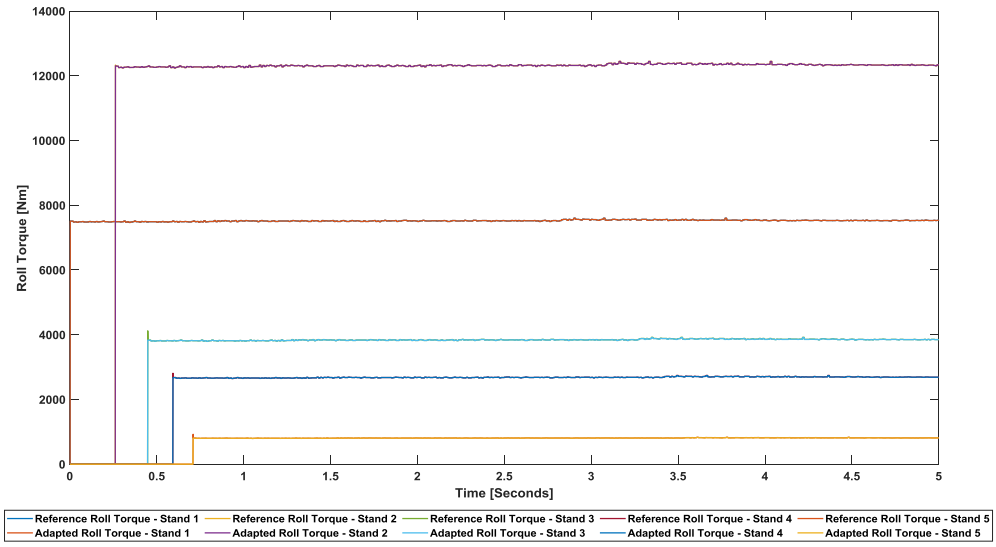

Figure 9. Adapted roll torque using DDPG roll torque agents.

VI. Roll Gap Control

The output material cross-sectional area and tolerances are highly influenced by many factors, e.g. roll gab of the mill stand, interstand tension, material temperature, etc. An experienced operator is often used in real plants to control the end product cross-sectional area and tolerances based on optimized setting values. The setting values are selected based on trial, error and the experience of the operators and engineers of this plant, which makes them vulnerable to being inaccurate and unreliable. Pass scheduling is used to determine the optimum number of passes to obtain a product with specific dimensions [29]. It takes into consideration the desired reduction percentage in each pass and the shape and size of the roll pass. Nevertheless, the method is not accurate as the setting values are obtained based on trial and error experience.

To reduce the downtime and improve product quality and tolerances, we propose an intelligent roll gab control using DDPG algorithm with actor-critic network configuration. The roll gap agent provides five control actions based on the observed states "20 states" which are used to adapt the roll gap setpoint value of each mill stand, see Fig. 10. The observed states are the roll force, roll torque, output temperature, and the cross-sectional area error for stands 1 to 5. The first three observations for all stands are available in the real plant dataset. In the process under study, however, there is no measurement available for the material cross-sectional area between the mill stands. Two sensors are only available which are placed at the beginning and the end of the mill stand block. To overcome this problem, the desired reduction percentage of the material cross-sectional area in each mill stand is used. The percent reduction of area for each mill stand can be easily obtained from the roll gap setting values available in the dataset and the roll gap model using (10), where A_0 is the original cross-sectional area of the material and A_r is the material cross-sectional after reduction. With the percent reduction of area, the output cross-sectional area error can be distributed to the five mill stands.

$$\text{Percent reduction of area} = \frac{A_o - A_r}{A_o}. \quad (10)$$

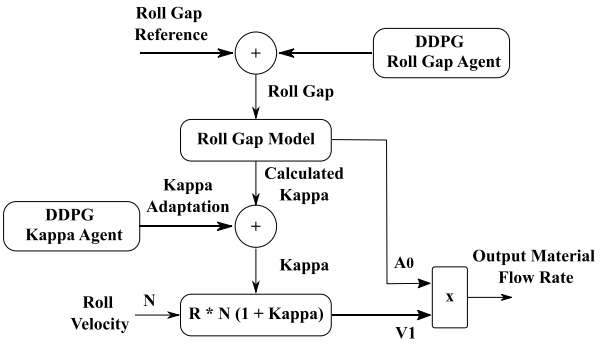

Figure 10. DDPG roll Gap Controller.

As mentioned earlier the roll gab DDPG-Agent has an actor-critic network configuration. The actor-network takes an input of size 20X1 and consists of four fully connected layers with 100, 80, 60, and 5 units respectively, each is followed by an activation layer. The first three activation layers have ReLU activation and the last layer has Tanh activation. The critic network has two branches; namely State and Action branches. The state branch takes an input of size 20X1 and consists of two fully connected layers with 100, 80 units respectively and one ReLU activation layer. The action branch takes as an input the action taken by the actor-network with size 5X1 which is then followed by a fully connected layer with 80 units. The output of both branches, i.e. State and Action branches are then concatenated and the output is given as an input to ReLU activation layer which is followed then by a fully connected layer with one unit. As the state and action spaces are continuous, a continuous reward function (7) is used.

The agents are configured to collect the maximum reward, i.e. long-term reward over an episode by setting the discount factor to 0.99. The actor and critic networks are trained from the random samples stored in the experience replay buffer. The experience replay buffer size is set to 1000000. And the mini-batch size is set to 256. The target smooth factor (τ) to transfer the weights to the target network is 0.001. The exploration factor of the agent is determined using the noise factor which is set to 0.05. Each episode starts when the material enters the last mill stand and terminates when the agent has generated actions for 1 second (1000 steps). The training is stopped when the agent achieves an average of 50 episodes' reward of 200 or above. Further, the agent is trained for an Area in the range of 38.5 to 41.5 mm^2.

To evaluate the performance of the trained roll gap agent, the agent is deployed in simulation. Fig. 11 illustrates the roll gap controller response for a setpoint change. The roll gap controller achieved a steady-state error of 0.12. Fig. 12 depicts the system output cross-sectional area for multiple setpoint changes. Further, the roll gap controller is tested against disturbances and was able to suppress disturbances with a slight increase in the steady-state error. To evaluate the roll gap agent performance further, regret is calculated. The agent achieved a regret score of 0.17.

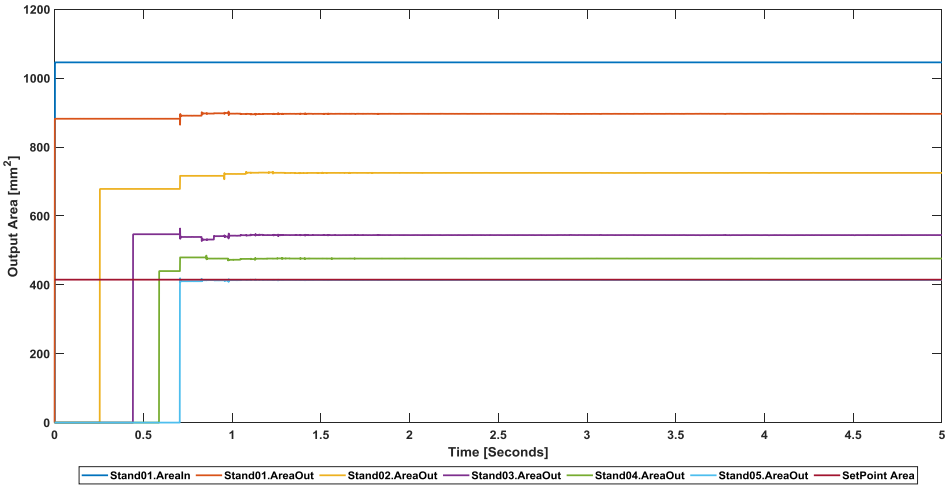

Figure 11. Roll gap controller response for setpoint change.

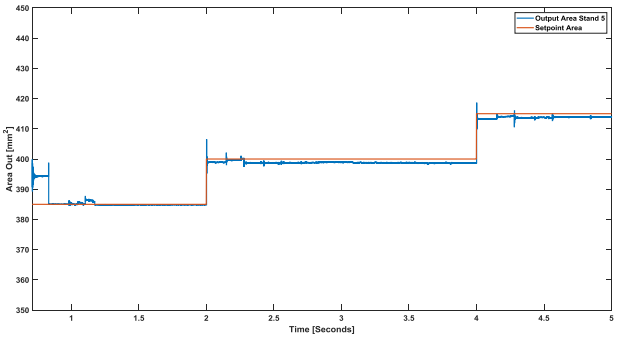

Figure 12. System output cross-sectional area for multiple setpoint changes.

VII. Conclusions

Although bar and wire hot rolling process is one of the oldest forming processes, it still lacks models that describe their underlying dynamics accurately. Most of the developed models assume a rectangle roll pass with an equal area for simplification, similar to flat rolling. These models, however, are uncertain owing to their inaccurate representation of process physics as well as unmodelled dynamics. Further, there is a high demand for improved control concepts to reduce downtime and improve product quality and tolerance. In this paper, gray-box model is adopted where reinforcement learning algorithms are employed to compensate for model inaccuracies using real plant measurement data. In addition, we presented an intelligent roll gab controller to reduce downtime and improve product quality and tolerance. The RL-agents are based on the Deep Deterministic Policy Gradients algorithm with the actor-critic structure, which can handle high-dimensional continuous action and state space. The developed DDPG agents showed high performance in model parameter identification achieving an overall identification error close to zero. Further, the roll gap controller was able to track setpoint changes and reject disturbances with a slight increase in the steady-state error.

In our future research, we intend to examine RL-agents with dynamic network architectures, e.g. Long Short-Term Memory (LSTM). Further, we will generalize the DDPG agents for different materials and mill stand block configurations.

Conflict of Interest

The authors declare that there is no conflict of interest.

Author Contributions

O. Gamal and M. I. P. Mohamed conducted the research, developed the idea, implemented the simulation and experimental work. C. G. Patel analyzed the data. H. Roth supervised the research. All authors contributed in writing the manuscript and had approved the final version.

ACKNOWLEDGMENT

This study is conducted under the EU-funded PIREF project (Prozessdiagnose und integrierte Reglung zur Effizienzsteigerung von Warmwalzstraßen für Stabstahl und Draht).

REFERENCES

[1] *Forming Handbook*, 2nd ed., Hanser, Munich, 2012, pp. 109-181.
[2] Y. Lee, *Rod and Bar Rolling: Theory and Applications*, 1st ed., New York: Marcel Dekker, 2004, ch. 2, pp. 9-28.
[3] J. B. Orr, *Rolls pass Design*, Sheffield, 1960.
[4] A. Hensel, P. I. Poluchin, and W. P. Poluchin, *Technology of Metal Forming, Ferrous and Nonferrous Materials*, Wiley-VCH, 1990.
[5] C. Overhagen and P. J. Mauk, "A new rolling model for three-roll rolling mills," *Key Engineering Materials*, pp. 879-886, 2014.
[6] H. Överstam and S. E. Lundberg, "A new approach to a model for the calculation of rolling force and rolling moment in wire and bar mills," in *Der Kalibreur*, 2008, pp. 65–74.
[7] R. Kawalla and W. Lehnert, "Rolling of bar and wire at the beginning of the 21st century," *MEFORM 2002*, Freiberg, 2002, pp. 1-26.
[8] H. Lippmann and O. Mahrenholtz, *Plastomechanics of Forming Metallic Materials*, Berlin: Springer-Verlag, 2013, pp. 529–551.
[9] O. Nelles, *Nonlinear System Identification: from Classical Approaches to Neural Networks and Fuzzy Models*, Berlin: Springer Science & Business Media, 2013, pp. 15–16.
[10] G. Lee, Z. Marinho, et.al., "Unsupervised learning for nonlinear piecewise smooth hybrid systems," arXiv, 2017.
[11] D. Liu, D. Wang, D. Zhao, Q. Wei, and N. Jin, "Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming," *IEEE Transactions on Automation Science and Engineering*, 2012, pp. 628-634.
[12] W. Yu and X. Li, "Some new results on system identification with dynamic neural networks," *IEEE Transactions on Neural Networks*, 2001, pp. 412-417.
[13] S. Chen, S. A. Billings, and P. M. Grant, "Non-linear system identification using neural networks," *International journal of control*, 1990, pp. 1191-1214.
[14] S. Ablameyko, "Neural networks for instrumentation, measurement, and related industrial applications," Netherlands: IOS Press, 2002, pp. 45–46.
[15] M. Jiang and Q. Jin, "Multivariable system identification method based on continuous action reinforcement learning automata," *Processes*, 2019, pp. 546.
[16] N. O. Lambert, D. S. Drew, et.al., "Low-level control of a quadrotor with deep model-based reinforcement learning," *IEEE Robotics and Automation Letters*, 2019, pp. 4224-4230.
[17] W. Yu, J. Tan, C. K. Liu, and G. Turk, "Preparing for the unknown: Learning a universal policy with online system identification," *arXiv*, 2017.
[18] R. Kamalapurkar, L. Andrews, P. Walters, and W. E. Dixon, "Model-based reinforcement learning for infinite-horizon approximate optimal tracking,"*IEEE Transactions on Neural Networks and Learning Systems*, 2017, pp. 753-758.
[19] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems Magazine*, 2012, pp. 76-105.
[20] A. M. Schaefer, D. Schneegass, V. Sterzing, and S. Udluft, "A neural reinforcement learning approach to gas turbine control," *International Joint Conference on Neural Networks*, 2007, pp. 1691-1696.
[21] A. M. Schaefer, S. Udluft, and H. Zimmermann, "A recurrent control neural network for data efficient reinforcement learning," *IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*, 2007, pp. 151-157.
[22] K. S. Hwang, S. W. Tan, and M. C. Tsai, "Reinforcement learning to adaptive control of nonlinear systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2003, pp. 514-521.
[23] M. Schäfer, O. Gamal, J. Wahrburg, and H. Roth, "Modeling, parameter estimation and validation of a bar and wire hot rolling process based on an example of a real rolling plant," *Automation 2019*, Baden-Baden, 2019, pp. 1035–1046.
[24] Q. Shi, H. K. Lam, B. Xiao, and S. H. Tsai, "Adaptive PID controller based on Q-learning algorithm," in *CAAI Transactions on Intelligence Technology*, 2018, pp. 235-244.
[25] T. P. Lillicrap, J. J. Hunt, et.al., "Continuous control with deep reinforcement learning," *arXiv*, 2015.
[26] G. D. O. Ramos, B. C. da Silva, and A. L. Bazzan, "Learning to minimise regret in route choice," in *Proc. of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 2017, pp. 846-855.
[27] L. Matignon, G. J. Laurent, and N. Le Fort-Piat, "Reward function and initial values: Better choices for accelerated goal-directed reinforcement learning," in *Proc. of International Conference on Artificial Neural Networks*, Berlin, 2006, pp. 840-849.
[28] S. C. Huang and J. Kim, "Control and System Identification of a Separated Flow," *Physics of Fluids*, New York, 2008, pp. 101509.
[29] S. Ray, *Principles and Applications of Metal Rolling*, U.K.: Cambridge University Press, 2016, ch. 3, pp. 104-157.

Omar Gamal is currently a Ph.D. student at the University of Siegen, Siegen, Germany. He received his B.Sc. degree in Mechatronics Engineering from the Higher Technological Institute, Ash Sharqiyah, Egypt. In 2017, he received his M.Sc. degree in Mechatronics Engineering from the University of Siegen, Siegen, Germany. His research interests include artificial intelligence, robotics, and control engineering.

Mohamed I. P. Mohamed received his B.Sc. degree in Mechanical Engineering from Periyar Maniammai Institute of Science and Technology, Thanjavur, Tamilnadu, India. In 2020, he received his M.Sc. degree in Mechatronics Engineering from the University of Siegen, Siegen, Germany. His research interests are artificial intelligence, autonomous driving, and high-performance computing.

Chirag G. Patel is currently pursuing his M.Sc. in Mechatronics engineering at the University of Siegen, Siegen, Germany. In 2016, he attained the B.Tech. in Mechanical Engineering from the Indus University, Ahmedabad, India. His research interests include artificial intelligence, computer vision, and human activity recognition.

Hubert Roth received the Dr.Ing. degree. He worked earlier in the space industry on attitude and orbit control systems for astronomical satellites. He is currently the Chair of the Control Systems Engineering at the University of Siegen, Siegen, Germany. He is also the Head of the Steinbeis Centre for Technology Transfer ARS. In both research and education, his emphasis is on control and sensor systems applied to mobile robots, spacecrafts, and swinging structures. He has a specific interest in virtual laboratories and tele-education in engineering.