Carbon emission reduction in railway maintenance using reinforcement learning

Sresakoolchai, Jessada; Kaewunruen, Sakdirat

DOI:
https://doi.org/10.1201/9781003323020

Life-Cycle of Structures and Infrastructure Systems – Biondini & Frangopol (Eds)

Carbon emission reduction in railway maintenance using reinforcement learning

J. Sresakoolchai & S. Kaewunruen
School of Civil Engineering, University of Birmingham, Birmingham, UK

ABSTRACT: At present, the effect of global warming becomes more severe due to carbon emissions. This negative effect attracts the interest of the international community to minimize it. The railway industry is an industry in which carbon emission cannot be neglected due to the sizes of projects, long operation and maintenance stages in the service life, including sources of energy that might not be clean in some areas. This study aims to minimize the negative effect of the railway industry on the maintenance aspect by using reinforcement learning to optimize maintenance activities. The maintenance in the railway industry is a complicated task and might not be optimized in terms of cost efficiency, serviceability, and environmental impact. The use of reinforcement learning can improve the overall efficiency of railway maintenance as it has been proven in many tasks in the railway industry and other industries. Data used to develop the machine learning model are based on field data collected during 2016-019. The length of the studies section is 30 km. Sources of data are from track geometry cars, maintenance reports, defect reports, and maintenance manuals of the sampled railway operator. The methodology used in the study is Proximal Policy Optimization (PPO). The results show that the use of reinforcement learning can reduce the carbon emission from railway maintenance activities by 48% which causes a significant amount of carbon emission while the railway defects are reduced by 68% which improved maintenance efficiency.

1 INTRODUCTION

Carbon emission is one of the causes of severe global warming and climate change nowadays. Therefore, carbon emission attracts more attention. The international community then tries to reduce it to minimize the negative effect on the environment. Although someones claim that railway transportation is one of the transportation modes that are environmentally friendly. However, it cannot be denied that many activities in the railway industry also create carbon emissions. For example, the construction of a railway project needs a lot of construction materials their production processes emit carbon or the sources of energy that are used to operate the railway may not totally clean. In addition, railway projects have long service life so the operation and maintenance phase also cause carbon. Therefore, minimizing carbon emissions in railway activities will be able to significantly minimize the negative effect on the environment.

This study aims to reduce carbon emissions from railway maintenance activities using reinforcement learning (RL). A technique used to develop a deep RL model is Proximal Policy Optimization (PPO). Data used to develop the RL model is collected from field data. The samples section is 30 km long. The duration of data collection is 2016-2019. Sources of data are from track geometry parameters collected using track geometry cars, track inspection reports showing track component defects, and maintenance reports showing performed maintenance activities during the interesting duration. Data from different sources are processed and combined together to feed into the RL model. Then, a maintenance manual of the sampled authority will be used to compare and define thresholds to create the customed RL

DOI: 10.1201/9781003323020-218

environment which will be used to define rewards for the RL's agent in the RL model training process. The expected contributions of the study are the developed RL model can be used by railway operators to plan railway maintenance activities more efficiently when carbon emission from maintenance activities is reduced as well as the number of railway defects. These will result in the improvement of an environmentally friendly railway system, and better serviceability, reliability, and safety in the railway system due to few defects occurring in railway networks. In addition, the developed RL model can be used to support decision-making for railway maintenance if railway operators are not ready to use a data-driven scheme in which decision-making completely relies on data and the RL model.

2 LITERATURE REVIEW

RL was first introduced in the early 1980s. It was developed to solve problems about how to react or what to do when we are in different situations (Sutton and Barto, 2018). The purposes of different reactions are to maximize rewards or outcomes from actions. As a representative, an agent is used and trained to know how to respond to different situations. The agent is not told how to do but it has to explore by itself how to maximize rewards at the latest stage. A challenge is each action not only affects the immediate rewards but also what happens next (stage). All of these are the important characteristic of RL that other types of machine learning techniques do not have and cannot resolve.

Examples of popular RL applications are chess and other game players, robots in industries, or pilot assistance in passenger cars (Sutton and Barto, 2018). Many methods are used to solve RL problems. One of the first methods is Tabular Solution Method which is the simplest. Then, other techniques were developed to be used with RL such as Markov Decision Processes, Dynamic Programming, or Monte Carlo Methods. During the last decade, many techniques have been developed to extend the potential of RL and other recent RL models have been developed and proven that the performance is better.

RL is one of the main categories of machine learning among supervised and unsupervised learning. RL becomes more popular because it has a capability that supervised and unsupervised have not. RL can react to specific environments and be trained to respond to environments that optimize rewards. Therefore, RL can be used to solve problems by finding the best responses or actions under specific situations. The interaction of the RL model is done through an agent over time or stage. The agent will use the process of trial and error to find the optimum policy under different situations.

In each timestep or stage, two components continuously interact which are the agent and the environment. The environment will provide information to the agent through stages and rewards. Then, the agent will react to the environment by taking action. After that, new stages and rewards will be provided by the environment to the agent again. This process will repeat until the end of the training. A simple flowchart of this process is shown in Figure 1. In every stage, a reward is provided to the agent based on how well the agent reacts to the environment. This will be used to find the best policy for the agent.

Reinforcement learning (RL) has been more popular in the last few years. It has been applied in many fields (Li, 2019) such as communication and networking (Luong et al., 2019), biology (Mahmud et al., 2018), electrical systems (Zhang et al., 2019), robotics (Kormushev et al., 2013), transportation (Abdulhai and Kattan, 2003), medical (Zhou et al., 2021), finance (Kolm and Ritter, 2020), or engineering (Andriotis and Papakonstantinou, 2019). Shaker et al. (2010) proposed a technique applying RL to assist in the landing of unmanned aerial vehicles. In that study, it was found that the RL agent could learn very fast and achieve the goal. Sun et al. (2017) applied RL to control and improve water use or agricultural irrigation. They found that RL could increase productivity while decreasing water use.

RL has been applied in the railway industry in different aspects. Šemrov et al. (2016) applied Q-Learning for rescheduling single-track trains. The action spaces of the agent consisted of two actions which were stop and go. Rewards were calculated based on delay. The model was tried with a three-station scenario. It was found that using RL could reduce delays

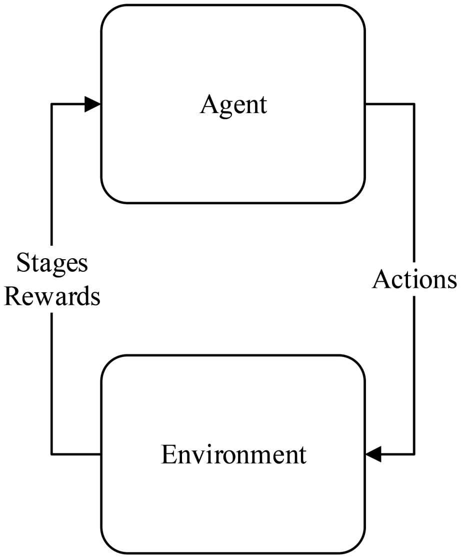

Figure 1. Simple flowchart of RL model.

in train schedules. This finding was also confirmed by Khadilkar (2018), Cui et al. (2016), Zhu et al. (2020b), and Cui et al. (2020). Other applications of RL in the railway industry were alignment optimization (Gao et al., 2022), power management (Xu and Ai, 2021, Zhu et al., 2020a, Deng et al., 2022), inspection (Zhong et al., 2021), and control (Wang et al., 2022).

For the environmental aspect, RL was used to manage the building energy (Yu et al., 2020) which showed that efficiency was improved by up to 30%. Cao and Wang (2022) applied RL to schedule the carbon emission of electric vehicles. They found that the use of RL could reduce carbon emissions by up to 20%.

From the literature review, there has never been a study applying RL for railway maintenance and carbon emission reduction. This research gap is interesting and is possible to be fulfilled using RL. Therefore, this study aims to apply RL to reduce carbon emissions generated from railway maintenance activities.

3 METHODOLOGY

3.1 *RL model*

This study applied PPO to develop the RL model. Referring to OpenAI (2017), PPO provides a better result than the state-of-the-art approaches while it is simple and easy to use and tune. PPO includes the supervised learning technique to implement a cost function. A concept is the RL model will try to minimize the cost function in each stage while keeping the deviation from the previous policy small. This creates less variance during the training process. As a result, the training will be smoother and prevent the agent from unreasonably reacting to the environment.

PPO applies the Actor-Critic scheme for the RL agent. The RL model comprises two parts which are the Deep Neural Nets of the Actor and the Critic. For the actor part, the model learns what action to take in specific situations. in other words, the agent will receive information from the environment. Information can be images or numerical data in the form of features likes supervised learning. For the critic part, the model will receive information about actions and observe the next stage to evaluate how well those actions are. The positive and negative outcomes will be evaluated and updated by the critic part. In conclusion, the actor part has the function to select actions and the critic part has the function to tell whether those actions are good or bad. From this process, the agent will have the ability to learn which actions are good. If the new policy is better than the current one, the agent will update its policy. If not, the agent will keep the current policy.

3.2 *Data characteristics and preparation*

Data are obtained from a selected section of the 30-km track during 2016-2019. Sources of data are track geometry measurements, defect inspection reports, and maintenance records.

Track geometry measurement contained data of superelevation, longitudinal level (10m chord), longitudinal level (20m chord), alignment (10m chord), alignment (20m chord), gauge, and twist (20m chord).

Defect inspection reports contain data on different track component defects. to simplify the defects, defects are categorized using types of track components consisting of ballast, fastener, rail, sleeper, and switch and crossing. Track geometry parameters and component defects will be used as states of the RL model. Therefore, there are 12 features used as states to determine whether track sections have defects. For component defects, it is straightforward. However, for track geometry parameters, defective sections will be considered using thresholds defined by the railway authority. If track geometry parameters exceed the thresholds, they will be considered defective. The thresholds are shown in Figure 2.

Maintenance records collect data on performed maintenance activities on each track section. There are seven maintenance activities which are tamping, rail grinding, ballast cleaning, sleeper replacement, rail replacement, fastening components replacement, and ballast unloading.

Every source of data is combined to process and analyze the characteristic of data. These data will be used to develop the custom RL environment. Maintenance activities can be considered seven binary action spaces. In each stage, different maintenance activities can be combined and it will create a lot of diversity in performed maintenance activities. However, the data sources are rich enough to be numerically processed to support this diverse action space. In other words, the size of the data is big enough to tell how much track geometry parameters will improve or deteriorate when any maintenance activities are performed. At the same time, how much the probability of each track component defect will increase or decrease when different maintenance activities are performed. More detail will be described in the following section.

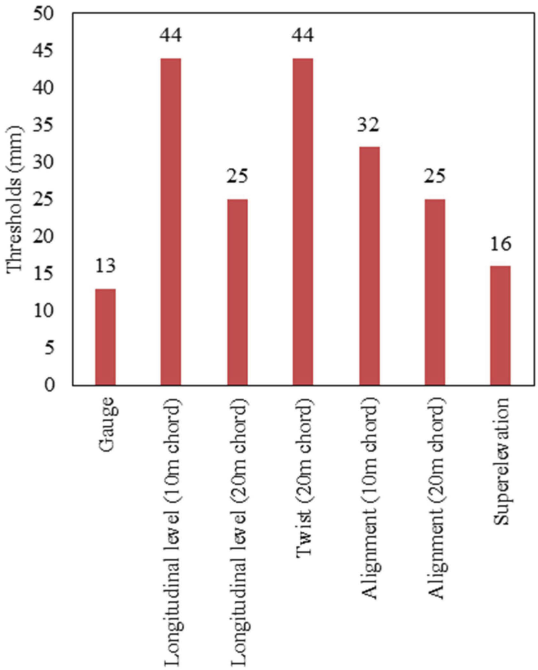

Figure 2. Track geometry thresholds.

3.3 *Problem description*

In the RL model, there are five components consisting of agent, environment, states, actions, and rewards.

In this study, the agent will learn how to perform maintenance activities according to 12 states which are seven track geometry parameters and five track component defects. There are five maintenance activities which are tamping, rail grinding, ballast cleaning, sleeper replacement, rail replacement, fastening components replacement, and ballast unloading that can be independently combined. Therefore, there are possible 128 combinations of maintenance activities. Different combinations of maintenance activities result in different improved and deteriorated track geometry parameters and the probability of track component defect occurrence that is obtained from the analysis of collected field data. The RL agent aims to minimize the maintenance activities to minimize the carbon emission while keeping tracks free from defects. Non-defective track means every track geometry parameter does not exceed defined thresholds and there is no track component defect.

The first stage of the RL model will have states or features as the field data. Then, the agent will randomly take actions that generate a new stage. This process will repeat 100 times which is the defined number for the agent to learn. Then, the reward will evaluate based on the actions. Rewards are defined by setting negative values of maintenance activities creating carbon emissions and penalties when tracks are defective. The penalties are set relatively high because maintenance aims to keep track in good conditions or defect-free. In a study done by Kaewunruen et al. (2021), they found that the maintenance activities of a 2.45 km track caused carbon emissions of 3,017 tCO2e. In this study, the length of a section is 1 foot or 0.3 m. Form the calculation, the average carbon emissions per maintenance activity in a section is 0.37 tCO2e. The agent is trained until the loss converges which demonstrates that the policy is optimized.

Then, results from field data and the RL model will be compared to evaluate the carbon emissions and the number of defects.

4 RESULTS AND DISCUSSION

The RL model is set to train for 100,000 epochs or training times. The RL model is developed using Python language. The technique used to develop the RL model is PPO as mentioned. The loss of the training is shown in Figure 3. It can be seen that after the model is trained for 30,000 epochs, the loss is equal or close to 0 which means the RL model is optimized and the training is done.

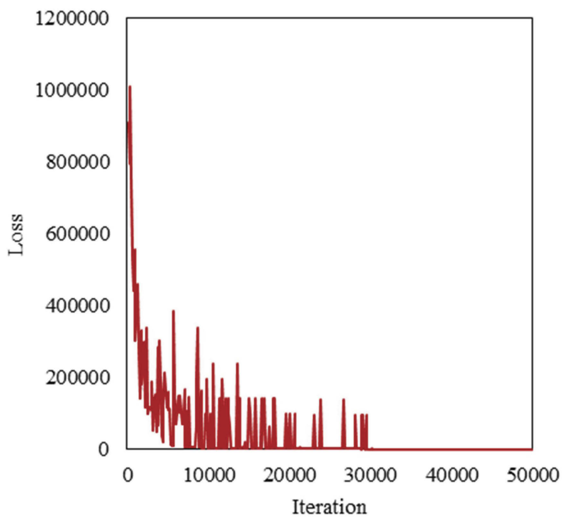

Figure 3. Training loss.

Figure 4 presents the differences between the number of performed maintenance activities and defects from field data and the RL model. From the field data, the number of performed maintenance activities and defects are about 963k and 520k respectively. However, the results from the RL model are 503k and 164k respectively. Therefore, the use of the RL model can reduce the number of performed maintenance activities by 48% and the number of defects by 68% which significantly improves the overall maintenance efficiency.

In terms of carbon emissions, assume that one maintenance activity causes a carbon emission of 0.37 tCO2e. Therefore, the maintenance activities from the field data will create a carbon emission of 356,404 tCO2e while the carbon emission from the maintenance when the RL model is applied is only 186,154 tCO2e which is only 52% of the current carbon emission. Therefore, it can be concluded that the developed RL model can significantly reduce carbon emissions from the maintenance activities in the railway industry which conforms to the aim of the study.

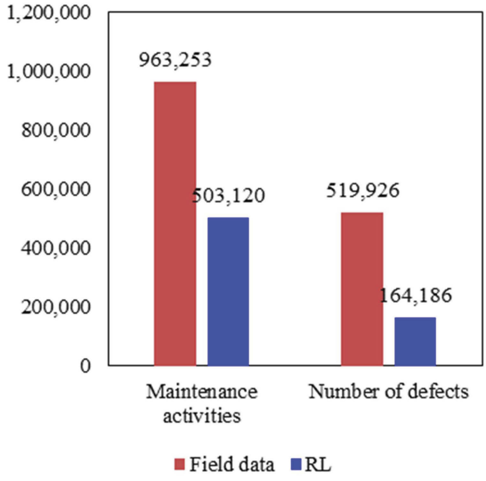

Figure 4. Differences in the number of performed maintenance activities and defects from field data and the RL model.

5 CONCLUSION

This study aims to apply the RL model to reduce carbon emissions from railway maintenance. The technique used to develop the RL model is PPO. States used to train the agent consist of 12 parameters from track geometry parameters and track component defects. Combinations of seven maintenance activities are used to create the action spaces for the Rl agent. Rewards are based on the number of performed maintenance activities and defects found in track sections. Therefore, the RL agent is trained to minimize the maintenance activities to reduce carbon emissions while keeping railway tracks defect-free. Data used to develop the RL model are from field data collected from the 30km track in 2016-2019.

The developed RL model can fulfill the aim of the study. It demonstrates the potential in reducing carbon emissions from railway maintenance by minimizing maintenance activities while limiting occurred defects. Compared to field data, the results from the RL model shows that it can reduce carbon emission by 48% and the number of defects by 68%.

Contributions of the study are the railway operators can use the concept from this study to develop their own RL model to support decision-making and plan the maintenance using their data. More states can be added to meet different requirements such as setting a limited budget for maintenance or time constraints to perform maintenance activities.

ACKNOWLEDGMENT

The authors also wish to thank the European Commission for the financial sponsorship of the H2020-RISE Project no.691135 "RISEN: Rail Infrastructure Systems Engineering Network", which enables a global research network that addresses the grand challenge of railway infrastructure resilience and advanced sensing in extreme environments (www.risen2rail.eu).

REFERENCES

Abdulhai, B. & Kattan, L. (2003) Reinforcement learning: Introduction to theory and potential for transport applications. *Canadian Journal of Civil Engineering*, 30, 981–991.

Andriotis, C. P. & Papakonstantinou, K. G. (2019) Managing engineering systems with large state and action spaces through deep reinforcement learning. *Reliability Engineering & System Safety*, 191, 106483.

Cao, Y. & Wang, Y. (2022) Smart Carbon Emission Scheduling for Electric Vehicles via Reinforcement Learning under Carbon Peak Target. *Sustainability*, 14, 12608.

Cui, L., Yuan, Z., Ming, Z. & Yang, S. (2020) Improving the congestion control performance for mobile networks in high-speed railway via deep reinforcement learning. *IEEE Transactions on Vehicular Technology*, 69, 5864–5875.

Cui, Y., Martin, U. & Zhao, W. (2016) Calibration of disturbance parameters in railway operational simulation based on reinforcement learning. *Journal of Rail Transport Planning & Management*, 6, 1–12.

Deng, K., Liu, Y., Hai, D., Peng, H., Löwenstein, L., Pischinger, S. & Hameyer, K. (2022) Deep reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles considering fuel cell aging. *Energy Conversion and Management*, 251, 115030.

Gao, T., Li, Z., Gao, Y., Schonfeld, P., Feng, X., Wang, Q. & He, Q. (2022) A deep reinforcement learning approach to mountain railway alignment optimization. *Computer-Aided Civil and Infrastructure Engineering*, 37, 73–92.

Kaewunruen, S., Sresakoolchai, J. & Lin, Y.-H. (2021) Digital twins for managing railway maintenance and resilience. *Open Research Europe*, 1, 91.

Khadilkar, H. (2018) A scalable reinforcement learning algorithm for scheduling railway lines. *IEEE Transactions on Intelligent Transportation Systems*, 20, 727–736.

Kolm, P. N. & Ritter, G. (2020) Modern perspectives on reinforcement learning in finance. *Modern Perspectives on Reinforcement Learning in Finance (September 6, 2019). The Journal of Machine Learning in Finance*, 1.

Kormushev, P., Calinon, S. & Caldwell, D. G. (2013) Reinforcement learning in robotics: Applications and real-world challenges. *Robotics*, 2, 122–148.

Li, Y. (2019) Reinforcement learning applications. *arXiv preprint arXiv:1908.06973*.

Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y.-C. & Kim, D. I. (2019) Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Communications Surveys & Tutorials*, 21, 3133–3174.

Mahmud, M., Kaiser, M. S., Hussain, A. & Vassanelli, S. (2018) Applications of deep learning and reinforcement learning to biological data. *IEEE transactions on neural networks and learning systems*, 29, 2063–2079.

Openai (2017) Proximal Policy Optimization.

Šemrov, D., Marsetič, R., Žura, M., Todorovski, L. & Srdic, A. (2016) Reinforcement learning approach for train rescheduling on a single-track railway. *Transportation Research Part B: Methodological*, 86, 250–267.

Shaker, M., Smith, M. N. R., Yue, S. & Duckett, T. (2010) Vision-based landing of a simulated unmanned aerial vehicle with fast reinforcement learning. IEEE.

Sun, L., Yang, Y., Hu, J., Porter, D., Marek, T. & Hillyer, C. (2017) Reinforcement learning control for water-efficient agricultural irrigation. IEEE.

Sutton, R. S. & Barto, A. G. (2018) *Reinforcement learning: An introduction*, MIT press.

Wang, H., Han, Z., Liu, Z. & Wu, Y. (2022) Deep Reinforcement Learning Based Active Pantograph Control Strategy in High-Speed Railway. *IEEE Transactions on Vehicular Technology*.

Xu, J. & Ai, B. (2021) Experience-driven power allocation using multi-agent deep reinforcement learning for millimeter-wave high-speed railway systems. *IEEE Transactions on Intelligent Transportation Systems*.

Yu, L., Qin, S., Zhang, M., Shen, C., Jiang, T. & Guan, X. (2020) Deep reinforcement learning for smart building energy management: A survey. *arXiv preprint arXiv:2008.05074*.

Zhang, Z., Zhang, D. & Qiu, R. C. (2019) Deep reinforcement learning for power system applications: An overview. *CSEE Journal of Power and Energy Systems*, 6, 213–225.

Zhong, J., Liu, Z., Wang, H., Liu, W., Yang, C., Han, Z. & Núñez, A. (2021) A looseness detection method for railway catenary fasteners based on reinforcement learning refined localization. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–13.
Zhou, S. K., Le, H. N., Luu, K., Nguyen, H. V. & Ayache, N. (2021) Deep reinforcement learning in medical imaging: A literature review. *Medical image analysis*, 73, 102193.
Zhu, F., Yang, Z., Lin, F. & Xin, Y. (2020a) Decentralized cooperative control of multiple energy storage systems in urban railway based on multiagent deep reinforcement learning. *IEEE Transactions on Power Electronics*, 35, 9368–9379.
Zhu, Y., Wang, H. & Goverde, R. M. P. (2020b) Reinforcement learning in railway timetable rescheduling. IEEE.