Imperfect Maintenance Optimization of Multi-State Rolling Stocks Based on Deep Reinforcement Learning

Chen Zhang
Department of Industrial Engineering
Tsinghua University
Beijing, China
chen-zha19@mails.tsinghua.edu.cn

Yan-Fu Li
Department of Industrial Engineering
Tsinghua University
Beijing, China
liyanfu@tsinghua.edu.cn

2021 3rd International Conference on System Reliability and Safety Engineering (SRSE) | 978-1-6654-0160-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/SRSE54209.2021.00051

***Abstract*—Developing an effective maintenance schedule for the rolling stocks has always been a critical issue of the railway companies. Currently, the Chinese railway companies still schedule the maintenance activities periodically according to the miles the rolling stocks traveled, which cause serious over-maintenance for the purpose of satisfying high reliability requirement. In this paper, we consider an imperfect maintenance optimization problem for multiple rolling stocks with stochastic maintenance time and operating conditions. The rolling stocks are modeled as the multi-state systems to characterize the degradation process. Meanwhile, the operating condition and the current location of the rolling stocks are also taken into consideration. The transition dynamics of the degradation process depends on the operating conditions. The optimization problem is formulated as a continuous-time Markov decision process and the objective is to maximize the total discounted reward related to the operating profit and the cost due to the maintenance, replacement and transportation in an infinite planning horizon. A deep reinforcement learning algorithm is developed to obtaining the optimal maintenance policy of the rolling stocks. A numerical experiment is given to demonstrate the advantage of the algorithm to improve the maintenance schedule of the rolling stocks.**

Keywords—imperfect maintenance, replacement, rolling stock, continuous-time Markov decision process, deep reinforcement learning

I. Introduction

Railway network plays an important role in the transportation infrastructure systems due to its high speed, super passenger capacity, and direct connections among cities. In China, the High-Speed Railway network serviced 1.73 billion passengers and had a fleet of 2600 rolling stocks in total in 2017 [1]. To guarantee the safety of the passengers, the reliability of the railway network is demonstrated during the system operations. The railway companies schedule complicated and detailed maintenance and inspection plans to satisfy the extremely high reliability requirement. In general, the maintenance of the railway network are usually divided into three major parts: 1) the maintenance of rolling stocks; 2) the maintenance of railway lines [2][3][4]; 3) the maintenance of the critical equipment in the rolling stocks [5][6]. Among these problems, the maintenance schedule of the rolling stocks shows great difficulties due to the movement characteristics of the rolling stocks as wheeled vehicles.

Currently, the maintenance plan of the rolling stocks is often based on regular inspection timetable which is produced manually by the manpower. The inspection activities are carried out mainly according to the miles the rolling stocks have been traveled. To avoid any occurrence of accidents threatening the lives of passengers, the traditional maintenance plan suffers serious over-maintenance situations. The over-maintenance scenario causes a waste of resources, and the maintenance cost is largely increased.

Therefore, developing a suitable maintenance plan is crucial to meet the reliability and economic requirement. A large body of the rolling stock literature uses mathematical programming to model the planning of operations and maintenance activities. Giacco, Carillo et al. [7] proposed a mixed-integer programming framework to solve the railway rostering planning covering services and maintenance works. Lai, Fan and Huang [8] developed a mixed-integer linear mathematical model of rolling stock assignment and maintenance plan considering depot capacity and circulation rules. Gu, Lam and Zinder [9] presented a nonlinear programming formulation for the rolling stock maintenance problem to get the optimal arrival dates to the maintenance center. Although maintenance activities are taken into consideration in these researches, the maintenance planning is always considered as a part of rolling stock operation schedules. The stochastic degradation process of the rolling stocks is not included as a basis for maintenance scheduling.

Some researches consider the stochastic properties involved in the rolling stock problems. Tönissen, Arts and Shen [10] introduced the uncertainty when locating maintenance facilities or changing line planning, fleet planning. Mira, Andrade and Gomes [11] considered the uncertain maintenance durations in their model. While these external factor influences the maintenance schedule, the inner degradation state is the most critical factor determining the maintenance time and maintenance level. It is worth mentioning that Erguido, Márquez et al. [12] presented a reliability-based maintenance modeling for the light rail, the dependency between the operating

978-1-6654-0160-9/21/$31.00 ©2021 IEEE
DOI 10.1109/SRSE54209.2021.00051

conditions and the degradation process is not considered in their model.

This paper presents an imperfect maintenance optimization model for multiple rolling stocks. The rolling stock is modeled as a multi-state system. The stochastic maintenance duration and the operating condition of the rolling stock are considered. The optimization problem to select the optimal maintenance level and the maintenance time is formulated as a continuous-time Markov decision process (CTMDP), in which the objective is to maximize the expected total discounted reward related to the operating profit, maintenance cost, replacement cost, and transportation cost. We present a proximal policy optimization (PPO) framework based on deep reinforcement learning to obtain the optimal maintenance policy for the rolling stocks.

The main contribution of this paper is in development of a novel maintenance optimization model for the rolling stocks based on the multi-state degradation model considering the operating conditions and location information. The second contribution is the implementation of deep reinforcement learning algorithm to overcome the computational complexity associated with the large action space.

The rest of this paper is divided in the following sections: Section I introduced the topic of the rolling stock maintenance and the related literature. Section II formulates the problem as a continuous-time Markov decision process model. Section III presents a deep reinforcement learning framework to solve the problem. In Section IV, the Makov decision process model is implemented. Finally, Section V presents the main conclusions and the directions for future research.

II. Problem Seetings and Formulation

In this section, the CTMDP formulation of our imperfect maintenance optimization model is presented. Then, the CTMDP model is transformed to an equivalent discrete-time Markov decision process (DTMDP) model for further solve the problem.

A. CTMDP Formulation

We consider the railway system consists of L rolling stocks in total. The state of each rolling stock is comprised of three parts: the degradation state, the operating state, and the location state.

Let the nonempty finite set $I \equiv \{0, 1, 2, \ldots\}$ denote the set of different operating conditions of the rolling stocks. The operating conditions of the rolling stock indicate its current operating environment of the rolling stock, the number of passengers it carries, etc. Since the operating conditions is stochastic, we make the assumption that the operation process evolves as a Continuous-Time Markov Chain (CTMC), $O_l \equiv \{O_l(t), t \geq 0\}, l = 1, 2, \ldots, L$, on the discrete state space I. The time duration between two jumps of the states is assumed to follow an exponential distribution which is already known. The mean time spent in different operating states are various. Thus, the transition rate between different states is not uniform. The rate of the operating mode $i \in I$ is ν_i. Let ν be a number that satisfies that

$$\nu_i \leq \nu, \ \ \forall\, i, \tag{1}$$

which is the upper bound of the transition rate in the CTMC of operating state.

We also assume that the operating state transition process of the different rolling stocks are independent. The probability that the operating state of the rolling stock l jumps from operating mode $i \in I$ to operating mode $j \in I$ after time $t \geq 0$ is defined as

$$P_{l,i,j} = \mathrm{P}\{\mathrm{O}_{\mathrm{l,i,j}}(t+s) \big| O_{l,i,j}(s)\}, \forall s \geq 0. \tag{2}$$

The rolling stocks are modeled as the multi-state systems, which subject to discrete degradation. The degradation process of the rolling stock is also modeled as a CTMC. Let $X_l \equiv \{X_l(t), t \geq 0\},\ l = 1,2,\ldots,L$ denote the degradation level of rolling stock l. Define the degradation state space $D \equiv \{1, 2, \ldots, F\}$, where state 1 to F is the degradation state from the perfect state to the failure state. Transitions only occur between two adjacent degradation levels.

Since the operating environment of the rolling stocks is exposed to nature, and the rolling stocks bear a huge pressure during the operating process, the degradation rate of the rolling stock highly depends on the current operating state. The degradation rate from state j to state $j+1$ at operating state i is denoted as

$$\lambda_{i,j} = 1, 2, \ldots, F - 1, \ \forall\, i \in I, \forall j \in D \backslash F. \tag{3}$$

During the operations of the rolling stocks, each rolling stock travels according to a predetermined schedule, which indicates the origin and destination of the rolling stock, and how long it will stay in the stations. Once a maintenance activity is needed, the rolling stock travels from one station to the maintenance bases. The maintenance bases are responsible for different maintenance activities. Since the rolling stock is moving on the railways and the locations of the maintenance bases are various, the travel time of the rolling stock varies according to the current locations and the maintenance activities. Besides, we take the stochastic characteristics of the maintenance time into consideration.

Let $Y_l \equiv \{Y_l(t), t \geq 0\}, l = 1, 2, \ldots, L$ denote the location of the rolling stock. Define the location state space $U = \{0, 1, 2, \ldots, N\}$. Location states 0 indicates that the railway is in the normal operation. Location 1 to N indicate that the rolling stock is arriving to, or leaving, or maintaining in the maintenance bases N to finish different levels of maintenance activities. Note that in reality, the first two low-level maintenance activities are situated in one maintenance base, and the rest three high-level maintenance activities are situated at another maintenance base. The requirement can be met by simply set the mean time at the states in the same maintenance base be the same.

We define the state of the rolling stock l as $S_l \equiv \{O_l, X_l, Y_l\}$. Then, the state of the entire CTMDP is $S \equiv \{s_1, s_2, \ldots, s_L\}$. We also define L-dimensional state vectors $O = \{O_1, O_2, \ldots, O_L\}$, $X = \{X_1, X_2, \ldots, X_L\}$, and $Y = \{Y_1, Y_2, \ldots, Y_L\}$. Then, the state is also a realization of the joint process $\mathrm{S} = \{O, X, Y\}$.

After observing the state of the rolling stocks at every time step, the decision maker takes a maintenance action from the action set. The maintenance action of rolling stock l is

$$a_l = \{0, 1, 2, 3, 4, 5, 6\}, l = 1, 2, \dots, L, \quad (4)$$

which can be one of the following:

0: Do nothing;

1 - 5: Different maintenance levels. Higher maintenance levels can bring the rolling stock to a healthy state with a lower degradation level;

6: Replacement.

The reward of the CTMDP triggers the rolling stocks keep a stable and reliable operations while reduces the maintenance cost to a low level. At each time step k, an operating profit p_l is given if the location of the rolling stock l is not in the maintenance bases, which is equivalent to $Y_l = 0$. If the imperfect maintenance action i is taken on rolling stock l, a maintenance cost $c_{l,i}^{PM}$ is incurred. Meanwhile, if the maintenance action i is taken place at the maintenance base n, the transportation cost related to the travel time between maintenance base n and the regular operating route of rolling stock l is $\mathrm{c}_{\mathrm{l,n}}^{TR}$. If a replacement happens on rolling stock l, a corrective maintenance cost occurs. We define the corrective maintenance cost as c_l^{CR} satisfying the relation that $c_l^{CR} \gg c_{l,i}^{PM} > \mathrm{c}_{\mathrm{l,n}}^{TR}$.

B. Equivelent DTMDP Formulation

We transform the above CTMDP $\hat{Y} = (S, A, \hat{r}, \hat{\mu}, \hat{p}, \beta)$ to an equivalent DTMDP by applying the theorem proposed by Serfozo 错误!未找到引用源。, where $\hat{r}$ is the reward function, $\hat{\mu}$ is the transition rate, $\hat{p}$ is the transition probability, and β is the discounted factor. Note that the CTMDP is a controlled Markov process with $\big(1 - \hat{p}(i, a, i)\big)\hat{\mu}(i, a) \le c < \infty$ for all $a \in A$ and $i \in S$. The CTMDP with nonuniform transition rates $\hat{Y}$ can be firstly transform to a CTMDP $Y = (S, A, r, \mu, p, \beta)$ with uniform transition rate c. With discounted rewards, the relation between reward function of the nonuniform CTMDP and the uniform CTMDP is

$$r(i, a) = \hat{r}(i, a)\big(\beta + \hat{\mu}(i, a)\big)/\,(\beta + c). \quad (5)$$

The relation between the transition probability of the nonuniform CTMDP and the uniform CTMDP is

$$p(i, a, j) = \begin{cases} \frac{\hat{\mu}(i,a)\hat{p}(i,a,j)}{c}, if\ i \neq j, \\ 1 - \frac{\hat{\mu}(i,a)(1-\hat{p}(i,a,i))}{c}, if\ i = j. \end{cases} \quad (6)$$

Based on the CTMDP with uniform transition rate, we have the DTMDP $X = (S, A, r, p, \frac{c}{\beta+c})$ be a controlled MDP with the same reward function and transition probability as that of Y defined in (4) and (5).

III. Deep Reinforcement Learning Algorithm

After deriving our equivalent DTMDP formulation, we adopt deep reinforcement learning as the solution framework, which has outperformed the classical meta-heuristics in many large-scale computational tasks. In particular, the Proximal Policy Optimization (PPO) algorithm is suitable for the problem with discrete action space as the model of ours. The PPO algorithm is developed from the Vanilla policy gradient algorithm and takes advantage of the actor-critic framework. The structures of the actor neural network and the critic neural network are shown in Fig. 1 and Fig. 2, respectively.

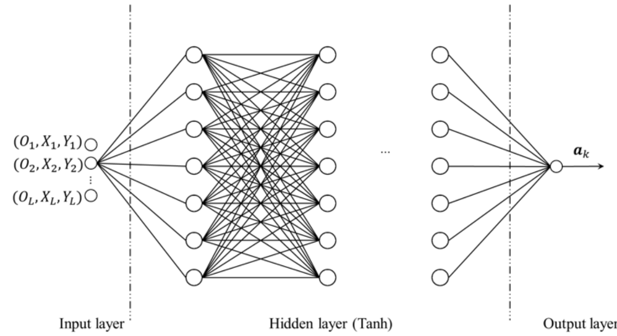

Fig. 1. The Structure of Actor Network.

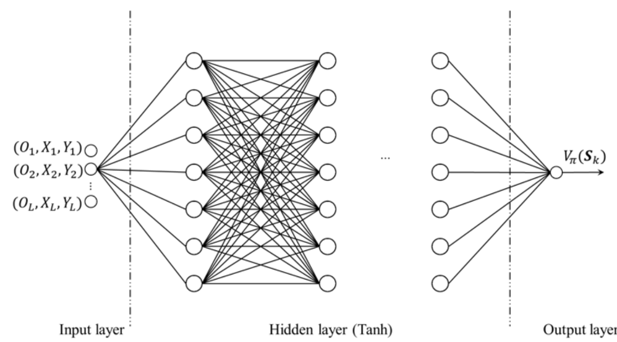

Fig. 2. The Structure of Critic Network.

The clipped surrogate objective function of PPO algorithm is given by

$$L^{CLIP}(\theta) = \hat{E}_t\big[\min\big(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1 - \epsilon, \epsilon)\hat{A}_t\big)\big], \quad (7)$$

where θ is the policy parameter, $\hat{E}_t$ is the empirical expectation over timesteps, r_t is probability ratio of the new and old policy, and $\hat{A}_t$ is the estimated advantage. Utilizing the probability ratio and truncating it between $[1 - \epsilon, 1 + \epsilon]$, PPO is simpler than the more complicated and hard-to-implement TRPO while maintaining the advantage of stability and fast convergence.

IV. Numerical Experiment

In this section, a numerical experiment is executed to show the procedure of the proposed methods and test the performance of the algorithm to search the optimal policy. Five rolling stocks are considered in our experiment. We consider two maintenance bases in accordance with the reality in China railway companies. The first maintenance base deals with the two lower-level

Identify applicable funding agency here. If none, delete this text box.

maintenance activities. The second maintenance base deal base the rest three higher-level maintenance activities.

Four operating states of the rolling stock $I = \{0, 1, 2, 3\}$ are considered where state 0 has the lowest number of passengers and state 3 has the highest number of passengers. The operating state transition rate and transition probability matrix is summarized in Table I and Table II.

TABLE I. OPERATING STATE TRANSITION RATE

Operating State	State			
	0	*1*	*2*	*3*
Transition Rate	240	150	420	90

TABLE II. OPERATING STATE TRANSITION PROBABILITY MATRIX

Operating State	State			
	0	*1*	*2*	*3*
0	0	0.3	0.7	0
1	0.5	0	0.1	0.4
2	0.2	0.6	0	0.2
3	0.1	0.1	0	0.8

Depending on the operating state, we consider eight degradation state with $D \equiv \{1, 2, \dots, 8\}$ in the case. The degradation state transition rate is related to the current operating state. The value of the state transition rate is summarized in Table III.

TABLE III. DEGRADATION STATE TRANSITION RATE

Degradation		State							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
a = 0	$O = 1$	0.5	0.5	1	1.5	2	2	2.5	3
	$O = 2$	0.5	1	2	3	4	4	5	6
	$O = 3$	0.5	2	4	6	8	8	10	12
	$O = 4$	0.5	3	6	9	12	12	15	18
a = 1	$O = 1$	680	680	680	680	680	680	680	680
	$O = 2$	640	640	640	640	640	640	640	640
	$O = 3$	610	610	610	610	610	610	610	610
	$O = 4$	570	570	570	570	570	570	570	570
a= 2	$O = 1$	540	540	540	540	540	540	540	540
	$O = 2$	510	510	510	510	510	510	510	510
	$O = 3$	420	420	420	420	420	420	420	420
	$O = 4$	400	400	400	400	400	400	400	400
a= 3	$O = 1$	390	390	390	390	390	390	390	390
	$O = 2$	380	380	380	380	380	380	380	380
	$O = 3$	350	350	350	350	350	350	350	350
	$O = 4$	320	320	320	320	320	320	320	320
a= 4	$O = 1$	240	240	240	240	240	240	240	240
	$O = 2$	235	235	235	235	235	235	235	235
	$O = 3$	230	230	230	230	230	230	230	230
	$O = 4$	220	220	220	220	220	220	220	220
a= 5	$O = 1$	150	150	150	150	150	150	150	150
	$O = 2$	130	130	130	130	130	130	130	130
	$O = 3$	100	100	100	100	100	100	100	100
	$O = 4$	90	90	90	90	90	90	90	90
a= 6	$O = 1$	750	750	750	750	750	750	750	750
	$O = 2$	750	750	750	750	750	750	750	750
	$O = 3$	750	750	750	750	750	750	750	750
	$O = 4$	750	750	750	750	750	750	750	750

We take the operational profit $p_l = 150$ for all rolling stock $l = 1, \dots, L$. The cost parameters considered for imperfect maintenance and replacement is given in Table IV.

TABLE IV. COST PARAMETERS UNDER DIFFERENT MAINTENANCE ACTIONS

Action	Maintenance Cost
0	0
1	20
2	40
3	80
4	160
5	320
6	800

The transportation cost between the maintenance base and the route of the rolling stocks is given in Table V.

TABLE V. TRANSPORTATION COST BETWEEN THE MAINTENANCE BASE AND THE ROUTE OF THE ROLLING STOCKS

Rolling Stock	Maintenance base	
	1	*2*
1	3	6
2	2	9
3	5	1
4	9	7
5	12	11

With the above parameter settings, we construct the simulation environment for the deep reinforcement learning algorithm. The discounted factor of the CTMDP is set to $\beta = 0.99$. The truncating parameter ϵ is set as 0.3. **错误!未找到引**

用源。then shows the expected discounted reward during the training process. The performance of the maintenance policy gradually increases during the training process.

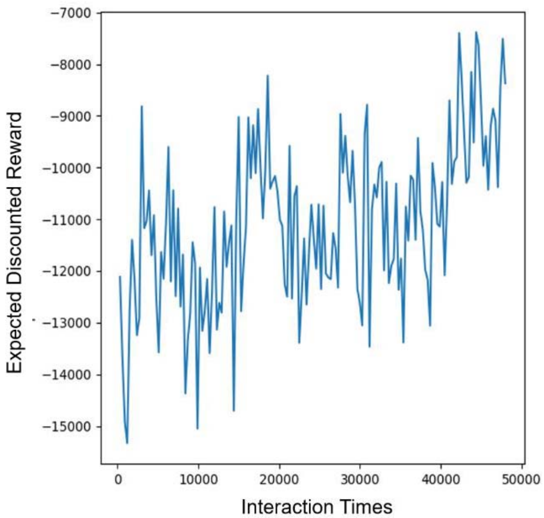

Fig. 3. Computational performance.

V. CONCLUSION

In this paper, we focus on the imperfect maintenance optimization of multi-state rolling stocks with considering the location and operating conditions. In order to reduce the long-run maintenance cost and guarantee the system reliability, the problem is formulated as a CTMDP problem. The CTMDP is further transformed into a DTMDP to decrease the difficulty when searching the optimal policy. We have utilized the PPO framework based on the deep reinforcement learning algorithm to obtain the maintenance policy. Numerical experiments are provided to present the procedure of the methods in detail and show the performance of the algorithm.

Several directions for further research exist. First, the detailed timetable rather than only the station information can be formulated into the model to get a more precise maintenance schedule. Second, the optimal policy of the CTMDP can be analyzed to get insights for accelerating the algorithm for solving the problem. Finally, the performance of the algorithm can be compared with the meta-heuristics method or other deep reinforcement learning approach to show the advantages of the algorithm.

ACKNOWLEDGMENT

This work was supported by National Key R&D Program of China under the Grants No. 2017YFF0209402.

REFERENCES

[1] Zhong, Q., Lusby, R. M., Larsen, J., Zhang, Y., & Peng, Q. (2019). Rolling stock scheduling with maintenance requirements at the Chinese High-Speed Railway. *Transportation Research Part B: Methodological*, *126*, 24-44.

[2] Macedo, R., Benmansour, R., Artiba, A., Mladenović, N., & Urošević, D. (2017). Scheduling preventive railway maintenance activities with resource constraints. *Electronic Notes in Discrete Mathematics*, *58*, 215-222.

[3] Nijland, F., Gkiotsalitis, K., & van Berkum, E. C. (2021). Improving railway maintenance schedules by considering hindrance and capacity constraints. *Transportation Research Part C: Emerging Technologies*, *126*, 103108.

[4] Sharma, S., Cui, Y., He, Q., Mohammadi, R., & Li, Z. (2018). Data-driven optimization of railway maintenance for track geometry. *Transportation Research Part C: Emerging Technologies*, *90*, 34-58.

[5] Zhang, D., Hu, H., Liu, Y., & Dai, L. (2014). Railway train wheel maintenance model and its application. *Transportation Research Record*, *2448*(1), 28-36.

[6] Lin, J., Pulido, J., & Asplund, M. (2015). Reliability analysis for preventive maintenance based on classical and Bayesian semi-parametric degradation approaches using locomotive wheel-sets as a case study. *Reliability Engineering & System Safety*, *134*, 143-156.

[7] Giacco, G. L., Carillo, D., D'Ariano, A., Pacciarelli, D., & Marín, Á. G. (2014). Short-term rail rolling stock rostering and maintenance scheduling. *Transportation Research Procedia*, *3*, 651-659.

[8] Lai, Y. C., Fan, D. C., & Huang, K. L. (2015). Optimizing rolling stock assignment and maintenance plan for passenger railway operations. *Computers & Industrial Engineering*, *85*, 284-295.

[9] Gu, H., Lam, H. C., & Zinder, Y. (2020). Planning rolling stock maintenance: Optimization of train arrival dates at a maintenance center. *Journal of Industrial & Management Optimization*.

[10] Tönissen, D. D., Arts, J. J., & Shen, Z. J. (2019). Maintenance location routing for rolling stock under line and fleet planning uncertainty. *Transportation Science*, *53*(5), 1252-1270.

[11] Mira, L., Andrade, A. R., & Gomes, M. C. (2020). Maintenance scheduling within rolling stock planning in railway operations under uncertain maintenance durations. *Journal of Rail Transport Planning & Management*, *14*, 100177.

[12] Erguido, A., Márquez, A. C., Castellano, E., Flores, J. L., & Fernández, J. G. (2020). Reliability-based advanced maintenance modelling to enhance rolling stock manufacturers' objectives. *Computers & Industrial Engineering*, *144*, 106436.

[13] Serfozo, R. F. (1979). An equivalence between continuous and discrete time Markov decision processes. *Operations Research*, 27(3), 616-620.