Traffic Energy Saving Control Based on Reinforcement Learning

Xiaoqing Zeng[1], Kaiyi Guo[1(✉)], Tengfei Yuan[2(✉)], Xiaoyuan Yue[1], Yizeng Wang[3], and Dongliang Feng[4]

[1] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, No. 4800 Cao'an Road, Shanghai 201804, China
Guo_Kaiyi@163.com
[2] SHU-UTS SILC Business School, Shanghai University, Shanghai 201800, China
yuantengfei@shu.edu.cn
[3] Shanghai University, Shanghai, China
[4] Shanghai Municipal Engineering Construction & Development Co., Ltd., Shanghai, China

Abstract. Train energy-saving operation control is a research hotspot in the field of urban rail transit energy-saving. By strengthening the perception ability and decision-making ability of the learning algorithm, this paper puts forward a new idea for the train energy saving control in urban rail transit under the condition of ensuring safety, comfort, real-time and punctuality. To be specific, the following work is done in this paper: (1) Study the related knowledge of train dynamics, establish the train traction model and the train running resistance model and complete the force analysis of the train motion process; (2) Study the knowledge related to energy consumption of train operation and establish the calculation model of energy consumption of trains within the interval; (3) Study the knowledge related to reinforcement learning algorithm, transform the train operation control process into Markov decision process, establish the three elements of reinforcement learning algorithm, and solve the train energy saving control problem by programming. Through simulation, the method proposed in this paper can reduce energy consumption by 13%–17% under the constraints of safety and punctuality.

Keywords: Urban rail transit · Energy saving · Train control · Reinforcement learning

1 Introduction

Compared with other modes of transportation, urban rail transit has the advantages of fast average speed, large passenger capacity, and high utilization rate. However, due to its high operating density and large passenger flow, the total resource consumption value of the urban rail system is also very huge. Among them, train traction energy consumption accounts for more than half. Therefore, studying the energy-saving control of train operation can promote a better green development of urban rail transit.

As shown in Fig. 1, this study minimizes the train operating environment as an optimization goal, takes operating safety, punctuality and other conditions as constraints,

X. Zeng et al. (Eds.): ITASC 2022, LNEE 1042, pp. 105–115, 2023.
https://doi.org/10.1007/978-981-99-2252-9_10

and uses different methods for model transformation. At present, the research methods on such issues are mainly divided into the following categories.

Zhu Jinling through the study of the principle of maximum value, introduced the limited speed constraint condition into the improvement model of the energy-saving driving strategy of the train, and the solution clarified the best control conditions [1]. Liang Zhicheng also used the principle of maximum value to further explore the problem of train handling under restricted speed constraints [2]. Wang Qingyuan used the relevant knowledge of maximum value to introduce the regenerative braking situation of multiple trains into the train control model, and clarified the best set of train control conditions [3]. Although the maximum value method can theoretically obtain the optimal numerical solution, it is precisely because of its too strong theoretical nature that the calculation is complicated.

Wang Pengling used the knowledge of dynamic programming to change the constraint conditions, and combined with the Gaussian pseudo-spectrum method to improve and optimize the train travel speed and distance curve [4].

Shi Hongguo discussed and studied the multi-objective problem of train operation. In order to improve the convergence speed and the performance of the output result, the genetic algorithm was introduced to improve the knowledge of variable length chromosomes, and then the problem was analyzed and solved and improved [5]. Liu Weianalyzed the improvement of multiple population genetic algorithms on the basis of predecessors, and introduced relevant knowledge of variable length real matrix coding, which was used in the improvement of the driving strategy of urban rail trains [6].

As one of the most concerned directions in the field of artificial intelligence in recent years, reinforcement learning provides continuous learning to control the behavior of agents by receiving high-dimensional perceptual input, providing a series of complex strategic decision and perception problems that are currently facing A new approach. Based on the relevant knowledge of the reinforcement learning algorithm, combined with the principle of maximum value, this paper uses the Markov decision-making model of the train running process in the interval, and uses the exploration and learning characteristics of the reinforcement learning to design a single-train energy-saving based on reinforcement learning. The control algorithm is used to solve the improvement strategy of train section driving, so as to obtain the optimal energy-saving strategy.

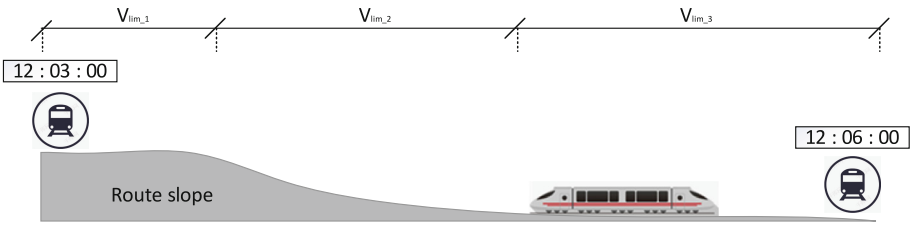

Fig. 1. Schematic scenario for this study

2 Analysis of Train Energy-Saving

When a train is running in a section, it will be affected by external forces such as train traction, train braking force, and train running resistance. Displayed equation is the calculation formula of the resultant external force.

$$F_C = F_T - F_B - F_R \tag{1}$$

FT is train traction (kN), FB is train braking force (kN), FR is train running resistance (kN). When the train runs in the section, it is affected by the traction, braking force, and running resistance of the train. According to the three laws of Newtonian mechanics, displayed equation is construct the kinematic equation of train.

$$\frac{dt}{ds} = \frac{1}{v} \tag{2}$$

$$\frac{dv}{ds} = \frac{\mu_t F_T - \eta_t F_B - F_R}{mv} \tag{3}$$

μ_t is train traction utilization coefficient (ratio of current traction force to maximum traction force, range [0,1]). η_t is rain braking force utilization coefficient (ratio of current braking force to maximum braking force, range [0,1]). The train will generate traction energy consumption during the driving process. This article aims to minimize the energy consumption of train traction, displayed equation is the minimum energy consumption function.

$$minJ = \int_{s_0}^{s_T} [\mu_t F_T(v) - \eta_t F_B(v)] ds \tag{4}$$

When the train is running, it must comply with the corresponding operating rules. Displayed equations are restrictions.

$$\begin{gathered} v(x) \le v_{Lim}(x) \\ v(0) = 0 \\ v(s) = 0 \\ t(0) = 0 \\ t(s) - t(0) = T_s \end{gathered} \tag{5}$$

V(x) is the speed of the train at position x (km/h). $v_{Lim}(x)$ is the speed limit of the train at position x (km/h). v(0) is the train's running at the beginning Speed (km/h); v(s) is the running speed of the train at the end (km/h). t(0) is the corresponding time at the beginning of the train (s). t(s) is the corresponding time at the ending of the train (s). T_s is the planned running time of the train in the interval (s).

Based on the knowledge of the maximum value, displayed equation is the Hamiltonian function of the train operation strategy.

$$H = -\mu_t F_T(v) + \eta_t F_B(v) + \lambda_1 \frac{\mu_t F_T - \eta_t F_B - F_R}{mv} + \lambda_2 \frac{1}{v} \tag{6}$$

λ_1 and λ_2 are the Lagrange multipliers. Displayed equations are the complementary relaxation factor M(x) to establish a regular equation.

$$\frac{d\lambda_1}{ds} = -\frac{\partial H}{\partial t} \tag{7}$$

$$\frac{d\lambda_2}{ds} = -\frac{\partial H}{\partial t} + \frac{dM(x)}{ds} \tag{8}$$

M(x) meets the constraint conditions of the displayed equations.

$$[v\ \ (x) - vLim\ \ (x)]\frac{dM(x)}{ds} = 0 \tag{9}$$

$$\frac{dM(x)}{ds} \geq 0 \tag{10}$$

Because there is no variable about time t in the Hamiltonian function, λ_1 is a constant. The Eq. (6) can be rewritten as follows equation.

$$H = (\beta - 1)\mu_t F_T(v) + (1 - \beta)\ \eta_t F_B(v) - F_R + \lambda_2 \frac{1}{v} \tag{11}$$

It can be seen from the above equation that if the function J wants to obtain the minimum value, the Hamiltonian function H needs to take the maximum value. Then, since the accompanying variables have multiple values, Table 1 is obtained.

Table 1. Optimal control value of train operation

β	μ_t	η_t
$\beta > 1$	1	0
$\beta = 1$	[0, 1]	[0, 1]
$1 > \beta > 0$	0	1
$\beta = 0$	0	1
$\beta < 0$	0	1

Depending on the value of the accompanying variable, the train operating condition may change (see Table 2). When $\beta > 1$, the value of the control variable corresponds to the maximum traction condition; when $\beta = 1$, the value of the control variable corresponds to any operating condition; when $\beta < 1$, the train corresponds to the maximum braking condition. Therefore, the energy-saving control strategy of train operation. Accelerate with the maximum traction force, then use the cruise mode or idle mode, and use the brake as little as possible during the non-braking operation phase. When the brake is required to stop, the maximum braking force is applied to the train. So, in the following algorithm design in this article, traction and braking are both adopted maximum traction and maximum braking, through the design of the algorithm, find the optimal idle point, in order to achieve as much energy-saving operation as possible.

Table 2. The value of the control variable of the train working condition

Train operating conditions	μ_t	η_t
Traction	(0,1]	0
Cruise	(0, 1)	(0,1)
Coasting	0	0
Brake	0	(0,1]

3 Research and Design of Reinforcement Learning Algorithm

In this paper, combined with the operation control mode of the train in the interval, the Markov decision process is selected as the environment model of the algorithm. Considering the delay time of information acquisition and the amount of calculation and storage of the algorithm, this paper sets the discrete time to 0.2 s. Then, the train is regarded as an agent, the control output of the train in each minute time period is regarded as the behavior set of the agent, and the speed and position of the train in each minute time period are regarded as the current state vector of the train.

Strategy: Strategy refers to the way an agent behaves at a given time or at a given stage, and directly determines the agent's actions or control decisions. In this paper, by modeling the operation process of the train section as a Markov decision process, the output set of the train controller is taken as the action set of the Markov decision process. Then, the action of the agent at stage i is expressed as the displayed equation.

$$u_{\mathrm{i}} = u_{i-1} + \Delta u_i \tag{12}$$

$\mathrm{u_i}$ is the train output of the i stage (kN). $\Delta \mathrm{u_i}$ is the train output change of the i stage (kN). Normally, the acceleration of the train in urban rail transit changes between $[-1, 1]\mathrm{m/s^2}$, but for passenger comfort considerations, the maximum traction acceleration of the train in this article is 0.6 $\mathrm{m/s^2}$, and the maximum braking The deceleration is -0.8 $\mathrm{m/s^2}$. Therefore, the change range of the acceleration change rate $\Delta \mathrm{u_i}$ is selected between $[-0.3, 0.3]\mathrm{m/s^2}$. In addition, another benefit of reducing the acceleration change rate range is that it can reduce the control variable range to a greater extent, thereby reducing the actual calculation data volume and data buffer volume of the algorithm, and improving the convergence calculation speed of the algorithm.

Reward Function: Reward function refers to the reward value generated by the agent in the process of moving to the next state due to an action taken in the current state. Because this article is to solve the problem of energy-saving operation control of trains, the energy consumption of trains in each small time period is regarded as a reward function. The specific calculation formula is as displayed equation.

$$U_i = \frac{1}{2} * \left| u_0 + \sum_{k=1}^{i} \Delta u_k \right| (v_i + v_{i-1}) \Delta t \tag{13}$$

u_0 is the output of the train starting (kN). v_i is the final speed (m/s) of the i stage of the train. Δt is the time of each discrete time period (0.2 s). Under normal circumstances, the calculation of energy consumption of trains is an integral process of force with respect to distance. But in this article, because the train operation control is broken down into minute time units. Therefore, in a discrete time period, it is approximately considered that the amount of change in force can be ignored, and the calculation of distance can be approximated as a distance calculation method in a straight-line process with uniform acceleration.

Value Function: The value function refers to the weighting and expectation of the reward function, which represents the overall return expectation of the entire control process of the strategy adopted by the agent in the current state. In this paper, the value function is defined as the displayed equation.

$$Q(X,\ \Delta u) = \frac{1}{k}\sum_{i=1}^{m-1}\gamma U_i \tag{14}$$

γ is discount factor (used to measure immediate repayment and long-term repayment, the range is (0,1]). After defining the state, strategy, reward function and value function above, the value function update formula (15) of reinforcement learning can be derived.

$$Q^{\Pi}(X,\ \Pi'(X)) = \frac{\varepsilon}{|A(X)|}\sum_{\Delta u} Q^{\Pi}(X,\ \Delta u) + (1-\varepsilon)\min_{\Delta u} Q^{\Pi}(X,\ \Delta u) \tag{15}$$

ε is greedy rate of greedy algorithm. A(X) is the number of action sets in state X.

Table 3 shows the process of the reinforcement learning algorithm to calculate a train speed-distance curve. Because this article needs to solve the problem of energy-saving optimization of train operation control, it is necessary to repeatedly calculate the process to obtain multiple train speed-distance curves, and then make selections to optimize the energy consumption of trains traveling in sections.

4 Simulation and Verification

4.1 Case 1

Case 1 uses the line data of Shanghai Metro Line 17 to verify that the energy-saving algorithm proposed in this paper is effective and reliable in terms of safety, punctuality and energy saving. In this case, the line data from Metro Line 17 from Jiasong Middle Road Station to Xujingbeicheng Station will be used. The running distance between the two stations is 2660 m, the planned train running time of this section is 180 s, and the section information of this section is shown in Table 4.

By importing the relevant line data into the reinforcement learning algorithm for calculation, the most energy-efficient train speed-distance curve in this section is obtained, as shown in Fig. 2.

As shown in the figure, the optimal strategy derived by this algorithm: first perform the maximum traction acceleration, then repeatedly perform the process of idling and

Table 3. Reinforcement learning algorithm calculates energy saving strategy

Step 1: Initially clear all parameters, reset cycle i, speed, distance, status and other information;

Step 2: Control the start of the train with the maximum traction acceleration;

Step 3: Determine the current position of the train S_i>S-S_d, where S_d is the train decelerating at the current speed. The required distance. If yes, go to step 9, otherwise go to step 4;

Step 4: Calculate the corresponding reward function U_i according to the state of the train at the last moment and the selected action;

Step 5: Use the greedy algorithm to select the action in the current state, and adjust the output of the train according to the above formula;

Step 6: Determine whether the output selected by the greedy algorithm meets the safety of train operation and other restrictions. If it is satisfied, the output selected by the greedy algorithm is used as the final output; otherwise, the output is adjusted and the output that meets the conditions is used as the final output;

Step 7: Obtain the final output and the state at the next moment, and update the value function according to the value function update formula above;

Step 8: i=i+1, go to step 3;

Step 9: Apply the train with the maximum braking deceleration to make the train stop at the station.

Table 4. Case 1 Line parameter information table

interval/m	slope/%	Speed limit/m
0–200	0	55
200–300	0	80
300–930	– 6	80
930–1800	– 3.081	80
1800–1980	– 3.081	60
1980–2660	0	60

traction, and finally act on the train with the maximum deceleration. And it is not difficult to find that the speed of the train is always below the speed limit of the section during the entire period of the train running. Therefore, this algorithm meets the safety requirements.

Then, according to the actual energy consumption of the train and the running time of the section, it can be obtained from the existing signal system to further enhance the verification of the effectiveness of the algorithm, and refer to the conclusion data [7]. Make a comparison (see Table 5). It can be seen that in terms of travel time, the existing driving strategy is the fastest, followed by this article, and the reference strategy is the slowest. In terms of operating energy consumption, the existing driving strategy is the most energy-consuming, followed by this article, the strategy referred to is the

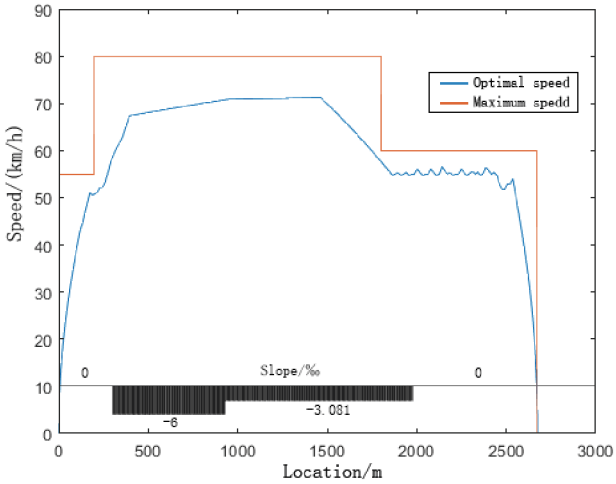

Fig. 2. Optimal control curve diagram based on reinforcement learning algorithm

best. Among them, compared with existing strategies, this strategy can reduce energy consumption by 17.5%. Compared with the energy-saving strategy in the reference, the strategy in this article has a slightly higher energy consumption, but the running time is closer to the standard time. In summary, the algorithm proposed in this paper meets the requirements of safety, punctuality and energy saving in train operation, and has a good energy saving effect.

Table 5. Comparison of three strategies

Driving control strategy	Time/s	Energy consumption/(MJ)
Existing ATO control strategy	179.1	187.21
Reference strategy	181.5	152.44
Strategy proposed in this article	180.7	154.39

4.2 Case 2

In order to further enhance the verification of the effectiveness of the algorithm, Case 2 uses the line data and other information of another reference[7] to compare the output results between different strategies. This case shows that the algorithm in this paper can model in different intervals. Validity in. In this case, the running distance of the section is 1354 m, and the planned running time of the section is 110 s. The slope, speed limit and curve radius information of the section are shown in Table 6.

By importing the relevant line data into the reinforcement learning algorithm for calculation, the most energy-efficient train speed-distance curve in this section is obtained,

Table 6. Case 2 Line parameter information table

interval/m	slope/%	Speed limit/m	Curve radius/m
0–120	0	55	0
120–304	0	80	0
304–684	1.8	80	19
684–1304	– 3.5	80	15
1304–1354	0	80	5

as shown in Fig. 3. It can be seen from the figure that the optimal strategy derived by this algorithm is largely the same as the case: first, the maximum traction acceleration is performed, and then the process of idling and traction is repeated repeatedly, and finally the train is braked at the maximum deceleration. The subtle difference is that the second case starts from the section line 684 m from the downhill section, so the acceleration of the idling here is a positive number, so the train has been idling in this section of the road.

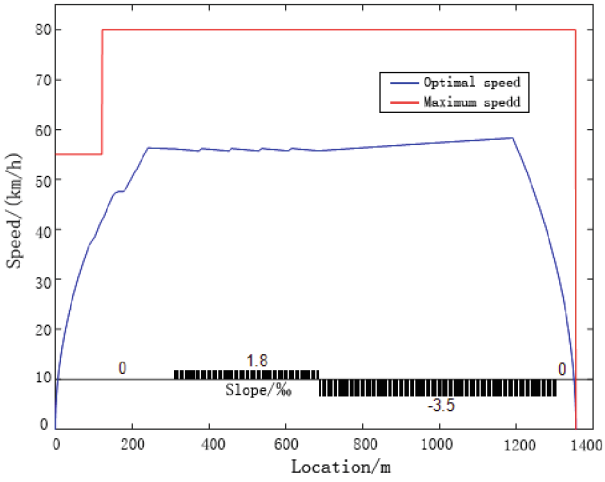

Fig. 3. Optimal control curve diagram based on reinforcement learning algorithm

Then, use the conclusion data in reference [9] to compare (see Table 7). Through comparison, it can be seen that in terms of running time, the existing driving strategy is the slowest, the algorithm in this paper is the second, and the strategy in reference to the literature is the fastest. From the perspective of the error time, the error time of this paper is only 0.2 s, which is 0.6 s lower than the error time of the existing strategy and is within the acceptable range. In terms of operating energy consumption, the existing driving strategy is the most energy-consuming, followed by references, and the strategy in this article is the best. Among them, the strategy in this article can reduce energy

consumption by about 13% compared with the existing strategy. Compared with the energy-saving strategy in the reference, the strategy in this article can further reduce the energy consumption by about 8% in this case. In summary, in this case, the algorithm in this paper also meets the requirements of safety, punctuality and energy saving, further verifying the effectiveness and reliability of the algorithm.

Table 7. Comparison of three strategies

Driving control strategy	Time/s	Energy consumption/(kw·h)
Existing strategy for case two	110.8	$3.06*10^7$
Reference strategy	110	$2.89*10^7$
Strategy proposed in this article	110.2	$2.66*10^7$

5 Summary

In this article, for the energy-saving operation of trains, the combination of reinforcement learning algorithm and train operation control is discussed, and the energy-saving optimization model and algorithm of urban rail single train based on reinforcement learning algorithm are proposed. According to the relevant data in the reference, the route and vehicle model are built, and the algorithm designed in this paper is used to solve the problem and verify the validity of the algorithm. Through the comparison of data results, this algorithm has achieved excellent results in both energy consumption and punctual safety.

Although the train energy-saving operation control system based on reinforcement learning designed in this paper has reached good requirements, there are also certain problems worthy of follow-up improvement, such as the establishment of a train multi-particle model.

Acknowledgements. The project is supported by Shanghai Science and Technology Committee Foundation (Number 19DZ1204202, 20dz1202903-0.1) and Shanghai Municipal Housing and Urban-Rural Construction Management Committee Foundation (Number JS-KY18R022-7).

References

1. Zhu, J., Li, H., Wang, Q., et al.: Optimization analysis of train energy saving control. China Railw. Sci. **29**(2), 104–108 (2008)
2. Liang, Z., Wang, Q., Lin, X.: Energy-efficient handling of electric multiple unit based on maximum principle. In: Proceedings of the 33rd Chinese Control Conference, pp. 3415–3422 (2014)
3. Wang, Q., Feng, X., Zhu, J., et al.: Simulation research on energy-saving optimal control of high-speed trains considering the utilization of regenerative braking energy. China Railw. Sci. **36**(1), 96–103 (2015)

4. Wang, P., Goverde, R.M.P.: Multiple-phase train trajectory optimization with signaling and operational constraints. Transp. Res. Part C **69**, 255–275 (2016)
5. Shi, H., Guo, H.: Multi-objective improved genetic algorithm for train operation simulation model. Railw. Transp. Econ. **30**(4), 79–82 (2008)
6. Liu, W., Li, Q., Guo, L., et al.: Research on optimization of energy-saving operation of urban rail trains based on multi-population genetic algorithm. J. Syst. Simul. **22**(04), 921–925 (2010)
7. Liu, L.: Research on Optimal Energy-Efficient Driving Evaluation of Metro Train Based on Flywheel Energy Storage. Tongji University (2020)
8. Miao, C., Wu, S., Zhou, Z., Zhang, W.: Research on energy-saving operation optimization of single train based on time discretization. J. Logist. Eng. Inst. **32**(03), 92–96 (2016)
9. Yin, J.: Research on integrated adjustment method of urban rail train operation based on approximate dynamic programming. Beijing Jiaotong University (2018)