Train Regenerative Braking Strategy Optimization Based on Reinforcement Learning

Xiaoqing Zeng[1], Liqun Liu[1], and Tengfei Yuan[2,3](✉)

[1] The Key Laboratory of Road and Traffic Engineering, School of Transportation Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai, China
[2] SHU-UTS SILC Business School, Shanghai University, Shanghai 201800, China
yuantengfei@shu.edu.cn
[3] Shanghai Engineering Research Center of Urban Infrastructure Renewal, Shanghai 200032, China

Abstract. With the increasement of urban rail transit operation density, the power consumption of metro system is also rising sharply. Meanwhile the proportion for urban rail transit of power consumption is increasing, so this problem needs more and more attention. In order to reduce the power consumption of rail transit, this research mainly focuses on the renewable energy utilization of train, which means that the train will make the best of the regenerative braking energy. For this purpose, the flywheel energy storage device is used as on-board device, then the regenerative braking strategy of the train is optimized based on reinforcement learning algorithm. Ultimately, the optimized train speed curve by the dynamic planning and Q-learning can achieve more than 5% energy recovery of the total energy consumption. The results show that this research can save the power consumption of rail transit by recycling the braking energy, which is of great significance for significance for energy saving and green transportation

Keywords: Regeneration energy · Dynamic planning · Reinforcement learning · Strategy optimization

1 Introduction

Urban rail transit is a kind of transportation mode with high reliability and large transportation capacity. With the improvement of the economic development of cities at all levels and the continuous improvement of engineering technology, subway, as a typical representative of urban rail transit, has been growing rapidly in the whole country. At present, China has ranked the first in the world in terms of operating mileage and passenger flow scale. More and more attention has been paid to the research of energy-saving operation of rail transit.

The main research goal of train energy-saving driving is to find the optimal speed curve when the train energy consumption is the lowest through the analysis of the speed curve of the train on the section under the condition of meeting the constraints of the operation environment. The optimal speed curve is mainly determined by the train force

X. Zeng et al. (Eds.): ITASC 2022, LNEE 1042, pp. 55–66, 2023.
https://doi.org/10.1007/978-981-99-2252-9_5

and the tangent sequence of working conditions, so there are many research results at home and abroad. Kunihiko [1] solved the train speed curve as a discrete bounded state variable problem by using the Pontryagin maximum principle. Howlett [2] et al. Proved the existence of single train energy-saving optimal operation method. Oshima [3] further applies fuzzy control theory to train operation control and improves the punctuality rate and stopping accuracy of automatic train driving system (ATO). Masafumi [4] et al. Introduced the consideration of the state of charge (SOC) of the energy storage device when studying the optimal control model, and comprehensively solved the optimal SOC control strategy. DOMínguez [5] et al. Studied the energy-saving control driving strategy when the train is equipped with ATO, and proposed a model solution based on multi-objective particle swarm optimization algorithm. Bao [6] and others put forward the simplified principle of train state space based on Pang's maximum principle. Liu [7] et al. used the value function approximation method to estimate the optimal value function, which improved the accuracy and operation efficiency of train operation strategy. Wu [8] et al. Combined with the control strategy of on-board energy storage device, studied the optimization of train speed curve. They used mixed integer linear programming (MILP) to solve the model.

Regenerative braking is also called feedback braking or regenerative braking in the field of rail transit. The biggest feature is to use the reversibility of the motor. When the train is braked, the traction motor reversely acts as the output energy of the generator, which is usually called regenerative energy. Renewable energy has great practical role and research value in many fields such as electric vehicles. In rail transit, the use of traction power supply system can achieve multi vehicle cooperation, thus reducing energy consumption at the system level. The principle of regenerative braking is shown in Fig. 1:

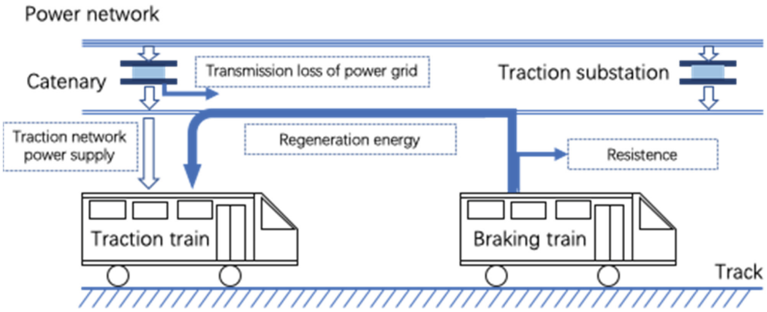

Fig. 1. Schematic of Direct feedback of train regenerative energy

The function of on-board energy storage device is to directly recover and store the regenerative energy generated by the train during braking, rather than feedback the traction network [9, 10]. Therefore, the on-board energy storage device can be used as an auxiliary power source to reduce the overall energy consumption of the traction power supply system under the condition of train traction. The schematic diagram of on-board energy storage device is shown in Fig. 2:

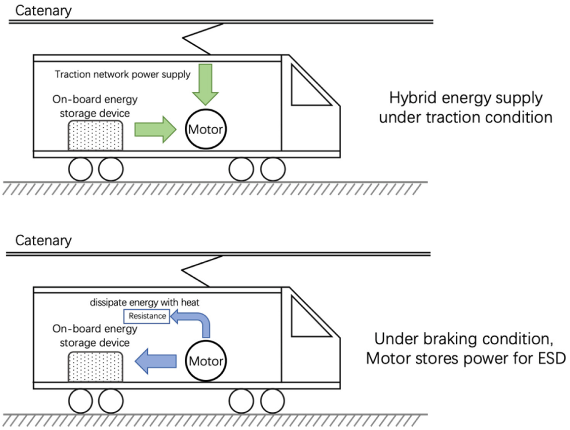
Catenary
Traction network power supply
On-board energy storage device
Motor
Hybrid energy supply under traction condition
Catenary
dissipate energy with heat
Resistance
On-board energy storage device
Motor
Under braking condition, Motor stores power for ESD

Fig. 2. Renewable energy utilization of train energy storage device

The purpose of this study is to analyze the optimization of the operation curve of the on-board energy storage train considering the recovery and utilization of regenerative braking energy, and to analyze and solve the evaluation of the optimal control curve of the train.

2 Model Formulation

When analyzing the forces on the running process of the railway train, it mainly needs to analyze the traction force, and running resistance of the train. It is very important for the calculation and optimization of the optimal speed curve to solve the stress state in different operation stages or working conditions. The train running resistance is generally divided into basic running resistance and additional running resistance, and the force situation is shown in Fig. 3:

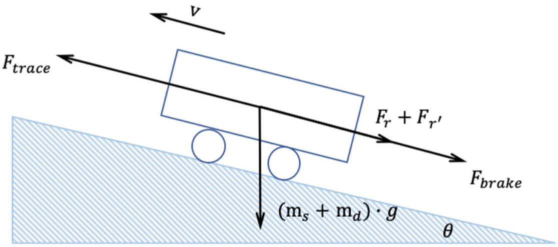

Fig. 3. Schematic of Single train force analysis

$$F_{trace}(v) = \begin{cases} F, & 0 \le v \le V_1 \\ \frac{P_0}{\mu_0 v + \varepsilon_0}, & V_1 \le v \le V_2 \\ \frac{P(v)}{\mu_1 v^2 + \mu_2 v + \varepsilon_1}, & V_2 \le v \le V_{max} \end{cases} \tag{1}$$

$$F_r = A + B \cdot v + C \cdot v^2 \tag{2}$$

$$F_{r\prime} = M \cdot g \cdot \sin\theta = M \cdot g \cdot i \cdot 1000 \tag{3}$$

where F is the constant traction output under the constant torque mode of the train, P_0 is the power of the train traction motor at constant power output, $P(v)$ is the output power of traction motor decreases with the increase of vehicle speed in the power reduction stage, μ_0, μ_0, μ_0 and ε_0, ε_1 are all represent fitting coefficient.

The basic resistance can be expressed by Davis Equation, and the main sources of additional resistance are: tunnel, turning radius and ramp. In this study, the ramp resistance is mainly considered. The resultant force on the train is indicated by $C(v)$:

$$C(v) = \begin{cases} \mu_f F_t(v) - \omega(v) - \omega\prime(x) \\ -(\omega(v) + \omega\prime(x)) \\ \mu_b F_b(v) - \omega(v) - \omega\prime(x) \end{cases} \tag{4}$$

$$a(v) = \frac{\mu C(v)}{(m_d + m_s)} \tag{5}$$

where μ is the coefficient of action of the resultant force, m_s, m_d represents the static load and dynamic load of the train respectively.

$$E = \sum_{k=0}^{n} C(v_k) \cdot s_k \tag{6}$$

The total energy consumption of the train in the operation section mainly depends on the stress of the train in the sub section. Where k is the number of subintervals, $k \in [0, n]$; s_k is the length of k subinterval.

$$E_{k,stroed} = T_k \cdot P_{f,r} = \frac{1}{2} J \left(\omega_k^2 - \omega_{k-1}^2 \right) \tag{7}$$

$$\omega_k^2 = \omega_{k-1}^2 + \frac{2E_{k,stored}}{J} \tag{8}$$

where $E_{k,stored}$ is energy storage of flywheel energy storage device in k subinterval; ω_k, ω_{k-1} represent the flywheel rotation speed in k and (k-1) subinterval; J is the moment of inertia of flywheel, kg·m^2; $P_{f,r}$ is rated power of flywheel energy storage device, w.

The objective function is shown in Eq. 9:

$$E = \min\left(\sum_{k=0}^{n} \mu F(v_k) \cdot l_{gap} - E_{k,stored} \right) \tag{9}$$

3 Solution Approach – DP

Dynamic programming (DP) is a theory first proposed by mathematician Richard Bellman in 1953 to solve multi-stage decision-making problems, rather than a specific algorithm or a specific mathematical model. Dynamic programming can effectively solve the problem with the property of optimality principle, which means that any decision subsequence contained in the decision sequence is always optimal, and satisfying the optimality principle can ensure that the discretized state has no aftereffect. So, the core of dynamic programming is to determine the Bellman equation according to the bellman optimality principle.

The problem can be discretized by evenly dividing the train operation intervals, and some sub intervals can be obtained [11]. Each interval corresponds to a driving state containing speed information, so that the problem of solving continuous speed curve can be transformed into a set of sub interval states. Suppose the length of the train operation section is l, and the operation section is divided into n sub sections. The length of each sub section is recorded as l_{gap}, and the set of all sub sections is recorded as S_k, thus (n + 1) states are divided. The state set corresponding to the subinterval can represent the vehicle speed and flywheel rotated speed at this position, which is shown in Eq. 10 (Fig. 4):

$$\begin{aligned} S_k &= \left\{s_{k,1}, s_{k,2}, \ldots, s_{k,n}\right\} \\ &= \left\{\left(x_{k,1}, v_{k,1}, \omega_{k,1}\right), \left(x_{k,2}, v_{k,2}, \omega_{k,2}\right), \ldots, \left(x_{k,n}, v_{k,n}, \omega_{k,n}\right)\right\} \end{aligned} \tag{10}$$

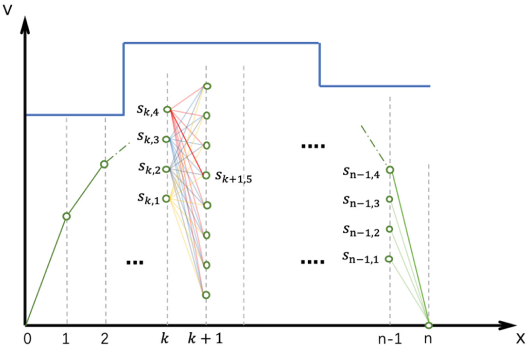

Fig. 4. Schematic diagram of dynamic planning state transfer

By traversing the state transition matrix and analyzing the time and energy consumption of each transition, the optimal utility function can be solved quickly in the process of dynamic planning, which provides the basis for action selection. The objective function of the optimal velocity curve is used to describe the state transition:

$$G\left(s_{k,v_i}, s_{k+1,v_j}\right) = \alpha\left(M_E - E^{k,i}_{k+1,j}\right) + \beta\left|T_{t,k} - T^{k,i}_{k+1,j}\right| + \gamma E_{k,stored} \tag{11}$$

where $G\left(s_{k,v_i}, s_{k+1,v_j}\right)$ is the function value of state transition; α, β and γ respectively represent the weight of energy consumption, time consumption and energy storage in the evaluation of transfer, which are greater than 0 and $\alpha + \beta + \gamma = 1$; M_E represents a constant for energy consumption, and as the reduced number of energy consumption $E_{k+1,i}^{k,1}$ corresponding to state transition, the lower the energy consumption, the better the effect; $T_{t,k}$ represents the reasonable time consumption of k stage in the interval, and the closer the actual time consumption is, the better the effect is.

$$G_k = \begin{bmatrix} G_{k,1} \\ G_{k,2} \\ \vdots \\ G_{k,n} \end{bmatrix} = \begin{bmatrix} G_{k_1}\left(s_{i_1}, s_{j_1}\right) & G_{k_1}\left(s_{i_1}, s_{j_1+1}\right) & \ldots & G_{k_1}\left(s_{i_1}, s_{j_1+m}\right) \\ G_{k_2}\left(s_{i_2}, s_{j_2}\right) & G_{k_2}\left(s_{i_2}, s_{j_2+1}\right) & \ldots & G_{k_2}\left(s_{i_2}, s_{j_2+m}\right) \\ \vdots & \vdots & \ddots & \vdots \\ G_{k_n}\left(s_{i_n}, s_{j_n}\right) & G_{k_n}\left(s_{i_n}, s_{j_n+1}\right) & \ldots & G_{k_n}\left(s_{i_n}, s_{j_n+m}\right) \end{bmatrix} \tag{12}$$

G_k in matrix G represents the result of the effect evaluation function of the transition from state s_i in phase k to s_j in phase k + 1. By this analogy, we can get the effect evaluation matrix of each stage in the whole process of the interval. According to the principle of Bellman optimality, we can get the optimal index function of the inverse solution of dynamic programming:

$$f_k\left(s_{k,v_j}\right) = \min\left\{ \begin{array}{c} G\left(s_{k,v_j}, s_{k+1,v_0}\right) + f_{k+1}^*\left(s_{k+1,v_0}\right) \\ G\left(s_{k,v_j}, s_{k+1,v_1}\right) + f_{k+1}^*\left(s_{k+1,v_1}\right) \\ \vdots \\ G\left(s_{k,v_j}, s_{k+1,v_m}\right) + f_{k+1}^*\left(s_{k+1,v_m}\right) \end{array} \right\} \tag{13}$$

$$f_k^*\left(s_{k,v_j}\right) = G^*\left(s_{k,v_j}, s_{k+1,v_i}\right) + f_{k+1}^*\left(s_{k+1,v_j}\right) \tag{14}$$

Equation 13 indicates that for the optimal index of state $s_{\mathrm{k,v_j}}$ in phase k, it is necessary to traverse all the states that can be converted to s_{k,v_j} in phase k + 1, and calculate the effect of these transformations. In Eq. 14, $f_k^*(s_{k,v_j})$ represents the optimal solution for the transition to $s_{\mathrm{k,v_j}}$ state, and $G^*(s_{k,v_j}, s_{k+1,v_i})$ represents the action corresponding to the optimal transition.

The above method is brought into the calculation example for verification, assuming that the information of operation section is shown in Table 1, and the information of train and vehicle is shown in Table 2:

Table 1. Section line information

Parameter	Symbol (unit)	Value
Line length	L (m)	1000
Maximum speed limit	V_{max}(km/h)	80
Standard operation time	T_t(s)	82
Floating range	Δ(s)	1
Stage gap	l_{gap}(m)	10

Table 2. Basic train information

Parameter	Symbol (unit)	Value
Static mass of train	M_s(t)	360
Dynamic quality of train	M_d(t)	386
Maximum braking acceleration	$a_{b,max}(m/s^2)$	1.1
Maximum traction acceleration	$a_{t,max}(m/s^2)$	1.1
Maximum traction	$F_{t,max}$(kN)	410
Davis Equation factors	A	10.079
	B	0
	C	0.001334

Table 3. Performance parameters of flywheel energy storage device

Parameter	Symbol (unit)	Value
Quality	M_f(t)	2
Minimum rotated speed	ω_{min}(rad/s)	280
Maximum rotated speed	ω_{max}(rad/s)	560
Rated rotated speed	ω_r(rad/s)	442.72
Moment of inertia	J $(kg \cdot m^2)$	180
Rated power	P_r(w)	3*105
Energy storage	E_r(J)	10.58*106

The basic parameters of flywheel mechanism of on-board energy storage device [12, 13] are shown in Table 3.

The optimized train speed curve can be obtained through the calculation as shown in the Fig. 5.

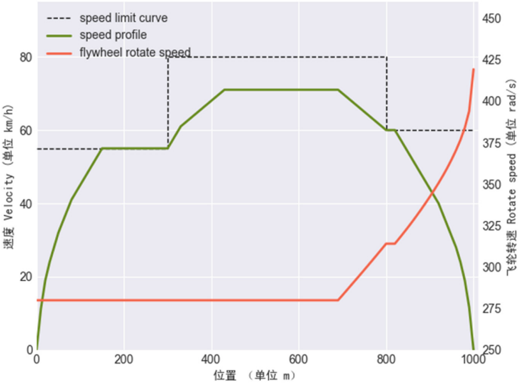

Fig. 5. Train speed curve calculated by dynamic planning

The test results show that the total energy consumption of the train running section is 166.82MJ, which can realize the recovery of the regenerative energy of the train about 8.4 mj, accounting for 5.25% of the total energy consumption. At the end of the section, the SOE of the flywheel energy storage device is 79.22%.

4 Solution Approach– Q-Learning

Reinforcement learning (RL) and deep learning, both of which have developed for many years, are the hot research categories in machine learning. Reinforcement learning is a kind of "environment" and "reward mechanism" established to evaluate the action and realize the optimal control of the decision-making process. The optimal control of urban rail train for the selection of train working conditions and the switching time of working conditions in the operation section conforms to the application direction of reinforcement learning. Q-learning is a kind of reinforcement learning without model. Compared with the model-based method, reinforcement learning without model does not need to bring into the model to solve the decision-making of the action subject [14], so as to plan in advance for the impact of the execution action.

Combined with Q table and R table, the updated Q value of action an in states can be calculated. The update of Q value is based on the Behrman function, just like the effect evaluation function in dynamic programming. The problem solved needs to meet no aftereffect. The Bellman equation on which the Q value is updated is shown in Eq. 15:

$$Q^{new}_{S,A} = Q_{S,A} + \alpha\left(R_{S,A} + \gamma * maxQ\prime(s\prime, a\prime) - Q_{S,A}\right) \tag{15}$$

where $\mathrm{Q}^{\mathrm{new}}_{\mathrm{S,A}}$ is the new Q-value after update; $Q_{S,A}$ is the Q-value in state S after performing action A; $R_{S,A}$ is immediate return on performing action A; $s^{'}$ is the state converted to after performing action A in state S; $maxQ^{'}\left(s^{'}, a^{'}\right)$ represents the maximum Q-value that can be obtained by all executable actions of state S in Q-table; α is learning efficiency, $0 < \alpha < 1$, the greater α is, the greater the weight of the existing Q value is; γ

is discount factor, $0 < \gamma < 1$, the greater the value of γ, the greater the weight of delay return.

To sum up, the pseudo code of Q-learning algorithm using $\in$ greedy search strategy to optimize the solution process of train speed curve can be expressed as follows:

Optimization of train speed curve based on Q-Learning

```
Start learning:
Import the speed curve solved by DP
Set Q-Learning parameters：$\alpha$、$\gamma$、$\epsilon$
Initialize Q-table，set all to 0 by default
Initialize R-table with return function，set -1 in R-table if action can't be implement
For i In Range(train_times):
    Initialize Q-table zero state
    For s In all_states: # States in ergodic intervals
        If sum(Q-table[s]) == 0 Or random.random() > $\epsilon$:
        # if Q-table[s] in initialization or random number is greater
            Randomly select one action implement in Q-value not equal to -1
        Else:
            Select the action with the largest Q-value in the table to execute
        Query R-table, update Q-table
        Initialize state $s'$
    End For
End For
return Q-table
Forward search Q-table, get the optimized speed curve
End of learning
```

The Q-value updating strategy in the solution process can be expressed as:

$$(s_k, a_i) = (1-\alpha)Q(s_k, a_i) + \alpha\big(R[s_k][a_i] + \gamma * \max\big\{Q\big(s\prime_{k+1}, a\prime_j\big)\big\}\big) j, i \in (0, n_a) \tag{16}$$

where $Q(s_k, a_i)$ is Q-value corresponding to a_i of state s_k; $R[s_k][a_i]$ is the return value corresponding action $a_j^{'}$ of state s_k in R-table; $\max\{Q\left(s_{k+1}^{'}, a_j^{'}\right)\}$ is delayed returns; n_a is the number of executable actions.

The calculation of the return function is shown in Eq. 17:

$$R\big(s_{k,a_i}\big) = (M_E - E(v_k, a_i)) + \alpha\big|T_{k,t} - T(v_k, a_i)\big| + \beta\big(T_{k,b} * P_{in}\big) \tag{17}$$

where M_E is the reduced constant of action energy consumption, then the greater the energy consumption and the lower the return; $E(v_k, a_i)$ represents the energy consumption of the action a_i performed in the k stage of the interval; $T_{k,t}$, $T(v_k, a_i)$ represents the time consumption in the k stage when the optimization is not carried out and the time consumption obtained by the execution of the action respectively. $T_{k,b}$ is the duration of brake application; P_{in} is the input power of energy storage device during train braking; α is the time consumption coefficient, which is mainly used to balance the order of magnitude of time consumption and energy consumption; β is the weight of renewable energy.

The main parameters related to Q-learning and the general settings of exploration strategies are shown in Table 4:

Table 4. Q-learning algorithm parameters

Algorithm parameter	Value
Learning rate α	0.1
Discount factor γ	0.9
Greedy coefficient ϵ	0.8
Learning times	3000

The comparison of the optimized speed curve is shown in the Fig. 6:

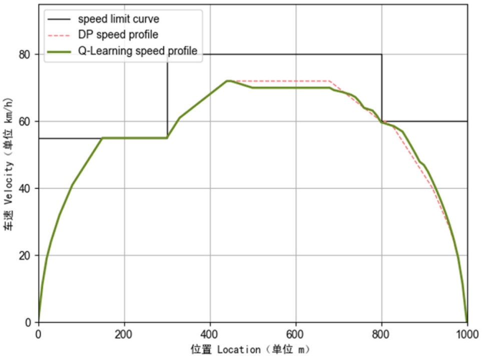

Fig. 6. Train speed curve calculated by Q-learning algorithm

In order to compare the train energy consumption, time consumption and regeneration energy recovery before and after optimization, Table 5 is established for analysis:

Through comparison, it can be found that: when using dynamic programming to solve speed curve, state discretization is based on uniform segmentation of speed interval δ, which helps to solve dynamic programming quickly and simplify state aggregation, and inevitably simplifies the speed after partial state transfer. In the uphill stage of the train (0−400 m interval in this case), if the train chooses the coasting condition, it will react on the speed instead of braking; but if the train chooses the coasting condition in the downhill stage (400−1000 m stage in this case), the speed will change slightly due to the short value of l_gap. In the forward search stage of dynamic planning, it will be

Table 5. Comparison of operation effect before and after Q-learning optimization

Algorithm	Total energy consumption MJ	Time consum *s*	Brake time consume *s*	E_{stored} MJ	SOE
DP	184.33	81.81	30.34	9.10	86.0%
Q-Learning	182.04	81.77	32.65	9.80	92.55%

due to the speed integration and conversion in the next stage the former is the same and recorded as cruise condition. This will lead to additional energy consumption of the train, and because the constant speed of IV in the working condition stage makes the speed curve tend to have greater braking force in the braking stage, which is reflected in the braking time of the train will not be conducive to the recovery of braking energy by the flywheel energy storage device.

Because the optimization based on Q-Learning is based on the response of the environment to the action, when establishing the return matrix in the forward search, the operation condition of the train is recorded according to the execution of the action. In this way, the train can use the terrain of the environment in the example to turn to coasting to reduce the running energy consumption, and according to the weight coefficient of the braking time in the call back function calculation, the better collection of the regenerative energy of the train can be realized. From the results in Table 5, it can be seen that through the optimization of Q-learning, the SOE of energy storage device in a single interval can increase the collection of renewable energy by 6.55%, and the total energy consumption in the interval decreases by 2.29mj, accounting for 1.24% of the total energy consumption of the original dynamic planning solution, while the difference between the operation time consumption in the interval and the standard time length is 0.232% and 0.28%, respectively, which are within the allowable error range.

5 Conclusions

Due to the proportion for urban rail transit of power consumption is increasing, so this uses the fly-wheel as the on-board energy storage device to save the braking energy. This method not only can effectively recycle the regenerative energy of the train, but also can reduce the overall energy consumption of the urban rail transit system. Based on the reasonable measure of selecting the fly-wheel as energy storage device, the dynamic planning and Q-learning algorithm is used to optimize train speed curve. The results show that the proposed method can achieve more than 5% energy recovery of the total energy consumption. Therefore, this research has some certain significance for significance for reducing energy consumption of rail transit. However, this research is only at the theoretical and simulation stage, so we need pay more attention to the fly-wheel using as the on-board device. In addition, the train regenerative braking strategy optimization algorithm needs further study.

Acknowledgments. The project is supported by Science and Technology Commission of Shanghai Municipality (20DZ2251900), as well as the Natural Science Foundation of Shanghai

(21ZR1423800). The work is also sponsored and supported by the Key Laboratory of Road and Traffic Engineering, Tongji University and Shanghai Engineering Research Center of Urban Infrastructure Renewal, Shanghai University. The authors are grateful for the reviewer of initial drafts for their helpful comments and suggestions.

References

1. Ichikawa, K.: Application of optimization theory for bounded state variable problems to the operation of train. Bull. JSME **11**(47), 857-865 (1968)
2. Howlett, P.: Existence of an optimal strategy for the control of a train. Sch. Math. Rep. **3** (1988)
3. Oshima, H., Yasunobu, S., Sekino, S.I.: Automatic train operation system based on predictive fuzzy control. In: Proceedings of the International Workshop on Artificial Intelligence for Industrial Applications, 1988. IEEE AI'88. IEEE, pp. 485–489 (1988)
4. Miyatake, M., Ko, H.: Optimization of train speed profile for minimum energy consumption. IEEJ Trans. Electr. Electron. Eng. **5**(3), 263–269 (2010)
5. Domínguez, M., Fernández-Cardador, A., Cucala, A.P., et al.: Multi objective particle swarm optimization algorithm for the design of efficient ATO speed profiles in metro lines. Eng. Appl. Artif. Intell. **29**, 43–53 (2014)
6. Bao, K., Lu, S., Xue, F., et al.: Optimization for train speed trajectory based on Pontryagin's maximum principle. In: International Conference on Intelligent Transportation Systems, pp. 1–6 (2017)
7. Liu, T., Xun, J., Yin, J., et al.: Optimal train control by approximate dynamic programming: comparison of three value function approximation methods. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2741–2746. IEEE (2018)
8. Wu, C., Lu, S., Xue, F., et al.: Optimization of speed profile and energy interaction at stations for a train vehicle with on-board energy storage device. In: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 1–6. IEEE (2018)
9. Shen, X.J., Chen, S., Zhang, Y.: Configure methodology of on-board super-capacitor array for recycling regenerative braking energy of URT vehicles. In: Industry Applications Society Meeting. IEEE, pp. 1678–1686 (2012)
10. Radcliffe, P., Wallace, J.S., Shu, L.H.: Stationary applications of energy storage technologies for transit systems. In: 2010 IEEE Electrical Power & Energy Conference, pp. 1–7. IEEE (2010)
11. Yeran, H.G., Lixing, Y., Tao, T., et al.: Train speed profile optimization with on-board energy storage devices: a dynamic programming-based approach. Comput. Ind. Eng. **126**, 149–164 (2018)
12. Gee, A.M., Dunn, R.W.: Analysis of trackside flywheel energy storage in light rail systems. IEEE Trans. Veh. Technol. **64**(9), 3858–3869 (2014)
13. Spiryagin, M., Wolfs, P., Szanto, F., et al.: Application of flywheel energy storage for heavy haul locomotives. Appl. Energy **157**, 607–618 (2015)
14. Lewis, F.L., Liu, D.: Reinforcement learning and approximate dynamic programming for feedback control. IEEE Circuits Syst. Mag. **9**(3), 32–50 (2015)