Vehicle emission control on road with temporal traffic information using deep reinforcement learning ⋆

Zhenyi Xu * **Yang Cao** ** **Yu Kang** *** **Zhenyi Zhao** ****

* *Department of Automation University of Science and Technology of China Hefei, 230026, China (e-mail: xuzhenyi@mail.ustc.edu.cn).*
** *Department of Automation University of Science and Technology of China Hefei, 230026, China, (e-mail: forrest@ustc.edu.cn)*
*** *Department of Automation University of Science and Technology of China Hefei, 230026, China. Institute of Advanced Technology, University of Science and Technology of China Hefei, 230088, China, (e-mail: kangduyu@ustc.edu.cn)*
**** *Department of Automation University of Science and Technology of China Hefei, 230026, China.(e-mail: zzy0025@mail.ustc.edu.cn)*

Abstract: The increased vehicle usage significantly aggravate the urban air pollution, which have great impact on the public health. Therefore, it is necessary to make proper traffic control policies and reduce traffic emissions. However, it is difficult to establish control strategies based on modeling methods, and carry out online control based on historical traffic information for the complex time-varying characteristics of emissions. In this paper, we present a deep reinforcement learning emission control strategy, which automatically learns the optimal traffic flow and speed limits to reduce traffic emission on the target road segment based on the temporal traffic information. The proposed approach is evaluated on real world vehicle emission data in Hefei. And the results demonstrate the effectiveness of the proposed approach against baseline methods.

Keywords: Urban air pollution, traffic emission control, deep reinforcement learning.

1. INTRODUCTION

Traffic emissions released by vehicles contain harmful substances, such as carbon monoxide (CO), carbon dioxide (CO_2), nitrogen oxides (NO_x), hydrocarbons (HC) and particulate matters (PM2.5), which have great impact on the public health. The increased vehicle usage significantly aggravate the urban air pollution. According to 'China Vehicle Environmental Management Annual Report (2018)', total CO emission is up to 33.273 million tons, HC is 4.071 million tons, NO_x is 5.743 million tons in 2017. In some megalopolis of China, such as Beijing, Shanghai, Shenzhen, the vehicle emission share rate of PM2.5 is approximately 13.5%-41%. Therefore, it is necessary to study traffic emission control to help relevant government departments make proper traffic control policies Pérez et al. (2000) and reasonable traffic planning Xu et al. (2020).

In recent years, there has been a lot of researches focusing on the issue of vehicle emission management. Most of the existing traffic emission control works can be classified into feedback control based approach and traffic management based approach.

The feedback control based method needs to establish macroscopic traffic flow model and microscopic traffic emission model for the controller to predict traffic emissions of traffic networks. Zhang and Zhu (2015) presented a discrete traffic flow model and designed a delay-feedback controller to suppress traffic jam and decrease traffic emission. Shu et al. (2013) proposed an integrated macroscopic traffic model to predict the traffic flow states and the emissions released by every vehicle at different operational conditions and designed the model predictive controller to reduce both travel delays and traffic emissions. Shuai et al. (2017) used multi-class macroscopic traffic flow and emission models for MPC for traffic networks to achieve a balanced trade-off between total time spent and total emissions. Uzunova et al. (2012) proposed non-integer robust control approach to analyse the speed and density variations due to perturbations on the road and to assure robust performances for the traffic velocity and evaluated pollution factor of CO2 emissions of the controlled and uncontrolled model. However, the macroscopic traffic flow model and microscopic traffic emission model are the simplification of the real world traffic condition, and the induced control strategy is definitely inaccurate.

⋆ This work was supported in part by the National Key R&D Program of China under Grant (2018AAA0100800, 2018YFE0106800, 2018YFC0213104), National Natural Science Foundation of China (61725304, 61673361), Major Special Science and Technology Project of Anhui, China (912198698036), as well as the Fundamental Research Funds for the Central Universities under Grant WK2380000001. (Corresponding author: Yu Kang)

The traffic management based approach mainly focus on the traffic speed and flow regulation to reduce the traffic emissions of the entire traffic networks. Panis et al. (2006) considers the effect of active speed management on traffic-induced emissions, suggesting vehicle acceleration and deceleration are important factors in determining traffic induced emission. Li et al. (2018) proposed an optimal dynamic credit charge scheme to redistribute the traffic flows to attain mobility and emission goals. Duell et al. (2014) presented a novel framework to estimate city-wide vehicle environmental effects that integrates a dynamic traffic assignment model with a novel application of a vehicle energy consumption model. Miles et al. (2018) proposed a decision support system (DSS) that uses an underlying traffic model to inform an atmospheric dispersion model. Bel et al. (2015) used quantile regression to evaluate whether speed management policies have been successful in promoting cleaner air, not only in terms of average pollutant levels but also during high and low pollution episodes. De Blasiis et al. (2014) presented an integrated simulation tool to analyse the effects of traffic flow conditions on the pollutants emissions. Panis et al. (2011) analysed different traffic types (urban versus highway traffic) with different modelling approaches (microscopic versus macroscopic) to examines the impact of speed management policies on emissions. Carslaw et al. (2010) used Generalized Additive Models to describe how emissions from individual vehicles vary depending on their driving conditions, taking account of variable interactions and time-lag effects, and quantified the impact that vehicle speed control has on-vehicle emissions of CO_2 by road type, fuel type and driver behaviour. Dijkema et al. (2008) studied the lowering of the maximum speed limit effects on reducing traffic related air pollution. The traffic management based approach designed the traffic control strategies based on the history traffic flow information, which is an offline strategy with time-lag infects.

And with the rapid development of traffic data collection and artificial intelligence techniques, some researches are shifting to use deep reinforcement learning (RL) approaches on traffic management. The RL learns directly from the interactions between states and actions through episodes, the agents take optimal actions according to the long-term accumulation of rewards. In this paper, we present a deep reinforcement learning emission control strategy, which automatically learns the optimal traffic flow and speed limits to reduce traffic emission on the target road segment based on the temporal traffic information. The main contributions of this study are as followings:

1) We proposed a road vehicle emission control reinforcement learning model to establish the relationship of emissions and traffic information. And a compound emission environment state space is designed to leverage the emission temporal dependencies. Moreover, to deal with the large state and action space, a deep return valuation network (DQN) is applied to estimate the optimal long-term value function.

2) The proposed approach is evaluated on real world vehicle emission data in Hefei. The results demonstrate the effectiveness of the proposed approach against baseline methods.

2. RELATED WORK

There are some published work on traffic control in transportation field, i.e. traffic lights control, traffic flow optimization, and vehicle speed regulation. Li et al. (2016) applied the deep reinforcement learning method to design signal timing plan by implicitly modeling the control actions and the change of system states. Nishi et al. (2018) developed a RL-based traffic signal control method that employs a graph convolutional neural network to cope with broader traffic demand. Walraven et al. (2016) solved the traffic optimization problem with reinforcement learning, where the traffic congestion on the highway is reduced. Li et al. (2017) incorporated the reinforcement learning technique in variable speed limit control strategies to reduce system travel time at freeway bottlenecks. Chao et al. (2013) proposed an indirect reinforcement learning model based on Dyna-Q architecture to manage incident-induced congestion for ramp control. A more general application of reinforcement learning in this domain can be found in work Cruciol et al. (2013), where different reward functions are investigated for decision-making in air traffic flow management with several stakeholders. Xu et al. (2019) proposed a deep spatiotemporal framework with multi-source urban data to predict the vehicle emissions in region scale. Yau et al. (2017) reviews various RL models and algorithms applied to traffic signal control in the aspects of the representations of the RL model (i.e., state, action, and reward), performance measures, and complexity to establish a foundation for further investigation in this research field.

3. OVERVIEW

3.1 Preliminary

The main notations used in this paper are shown in Table 1.

Table 1. Notations

Notation	Description
EF	An emission episode
TA	Traffic agent to conduction emission
tv_t	Traffic volume at time t (V/h)
ts_t	Traffic average speed at time t (km/h)
π	Traffic agent control policy

Traffic Agent: The traffic agent TA has the factors of traffic volume tv, and traffic average speed ts, which is the main body to interact with the emission environment.

Emission Episode: An emission episode EF is a period emission sequence in a day, in which the total vehicle emission we want to minimize. The emission consists of Fuel, CO, HC, and NO at each time interval, which can be defined as EF_t=$\{\text{Fuel}_t,\text{CO}_t,\text{HC}_t,\text{NO}_t\}$. And the sequences of emissions are all normalized respectively. In our problem, the length of emission episode is 24 hours, and the time interval is 1 hour.

3.2 Traffic data insight

Fig.1(a) and (b) show the daily speed and traffic volume distribution heatmap of the history traffic data. We find that in the morning rush hours, e.g. 10 a.m., the traffic pattern is with low speed and high traffic volume which easily contributes to high vehicle emissions, as can be seen in Fig.1(c).

3.3 Problem formulation

Problem: Given the history emission episodes EF, and traffic agent factors TA, design a control policy π to minimize the total emission amount in an episode while ensuring a normal traffic volume, as shown in Fig.2.

4. METHODOLOGY

Obtaining the history emission episodes, a reinforcement learning based model (EFRL) is proposed to minimize the total emission in the episode by learning traffic agent control policy.

4.1 Framework

As shown in Fig.3, our model consists of two phases, offline learning and online controlling.

Data Processing Given the original vehicle emission measurements and traffic related information from the emission remote sensing systems (emission environment), the EFRL structures the batched data into states s_t for generating the emission episodes.

Simulation In traffic emission variation process, there are interactions between vehicle emission and traffic information. Therefore, to train and evaluate a dynamic traffic emission model, a system simulator is required to simulate the system dynamics under repositions. During the simulation phase, the samples pool derived from real emission dataset to train our EFRL model. Specifically, the emission environment states, the traffic agent's coordination actions and immediate reward at each time step are used to train our EFRL DQN model for estimating the long-term value function, from which the coordination optimal traffic controlling policy can be inferred.

Online Controlling After offline learning phase, we obtain the trained EFRL DQN model, which returns optimal traffic control policy with traffic flow limits and vehicle speed limits suggestions to reduce vehicle emissions.

4.2 EFRL Model

A reinforcement learning model Sutton et al. (1998) is usually defined as a six tuples, i.e. (S,A,T,R,π,γ), where S is the state space of environment, A is the action space of agents, T is the transition probability which an agent took action a_t given state s_t will transit to the next state s_{t+1}, R represents the immediate reward after taking action under specific state, i.e. $r_t(s_t, a_t)$, π is a policy $S \times A \rightarrow \pi$, describing the probability to take an action given specific state, γ is a time discount parameter. The agent tries to select actions so that the sum of the discounted rewards it receives in the future is maximized, an action a_t has a long-term rewards G_t defined as Eq. 1

$$G_t = r_t + \gamma \cdot r_{t+1} + \gamma^2 \cdot r_{t+2} + \cdots + \gamma^k \cdot r_{t+k} \tag{1}$$

where $t+k$ is the final time step.

The objective of reinforcement learning is to learn an optimal policy π such that given state s_t , and taken action a_t, the agent is able to receive the maximum expected long-term discounted reward. The optimal long-term value function is defined as Eq. 2

$$Q^*(s_t, a_t) = \max_{\pi} \mathbb{E}_{\pi}(G_t | s_t, a_t, \pi) \tag{2}$$

Bellman equation is usually adopted to calculate Eq. 2, which is defined as

$$Q^*(s_t, a_t) = \mathbb{E}_{s_{t+1}}(r_t + \gamma \cdot \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1}) | s_t, a_t) \tag{3}$$

Given the optimal long-term value of each action by each state, we can get the optimal policy with Eq. 4

$$a_t^* = arg \max_{a_t} Q^*(s_t, a_t) \tag{4}$$

Based on the traditional reinforcement learning theory, we can formulate the emission control model as following concepts.

Observation Each time the traffic agent requires new control, it has a real-time observation of the current emission environment. The observation is defined to include the emission environment states, and the traffic agent states. For the emission environment states, which includes current emission states EF_t, and the observed emission $\widehat{EF}_t$ in the last period. For the traffic agent states, which includes the current observations of traffic volume $tv_{o,t}$, and traffic average speed $ts_{o,t}$. An observation can be denoted as $O_t = (EF_t, \widehat{EF}_t, tv_{o,t}, ts_{o,t})$.

Action The action of traffic agent is denoted as a vector describing the traffic volume limits and the traffic average speed limits, which is can be represented as $a_t = (tv_t, ts_t)$. For example, action $a_1 = (120, 40)$ denotes that the current traffic volume is limited to 120 V/h and the vehicle speed is limited to 40 km/h on the target road segment in the next hour.

State The state of the emission environment is defined as interleaved sequences of observations and actions combined with current time, which can be denoted as $s_t = (O_{t-k}, a_{t-k}, \cdots, O_{t-1}, a_{t-1}, O_t, t)$, k is the observation time lag. And k is set to 1 in this work.

Immediate reward To control the emission of an episode at low level while ensuring a quite traffic volume. The immediate reward is defined as a combination gain of emission and traffic volume r_t=$G_I(EF_t - EF_{t+1})$+$G_I(tv_{t+1} - tv_t)$ after taking action a_t under s_t and transiting to s_{t+1} in period (t,t+1]. And the definition of G_I is as

$$G_I(x) = \begin{cases} 1 & x > 0 \\ -1 & \text{otherwise} \end{cases} \tag{5}$$

Optimal value network Defining the EFRL model, we can design DQN to estimate the optimal long-term value function (Eq. 3), i.e. $Q^*(S, A, \theta) : S \times A \rightarrow Q^*$, where θ are the learning parameters. Our EFRL model is optimized by minimizing the Bellman equation square error, the objective function is formulated as

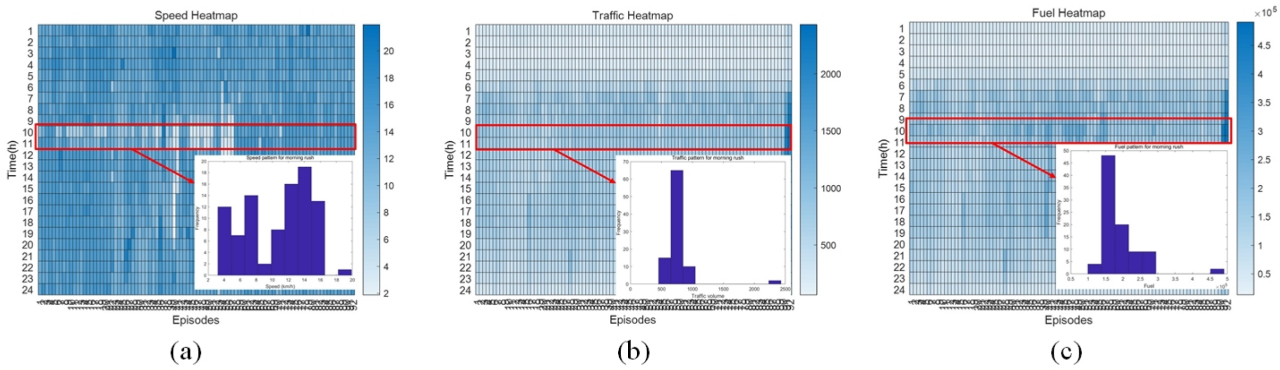

Fig. 1. Daily traffic and emission distribution: (a) low speed in the morning rush hours; (b) high traffic in the morning rush hours; (c) high emissions in the morning rush hours.

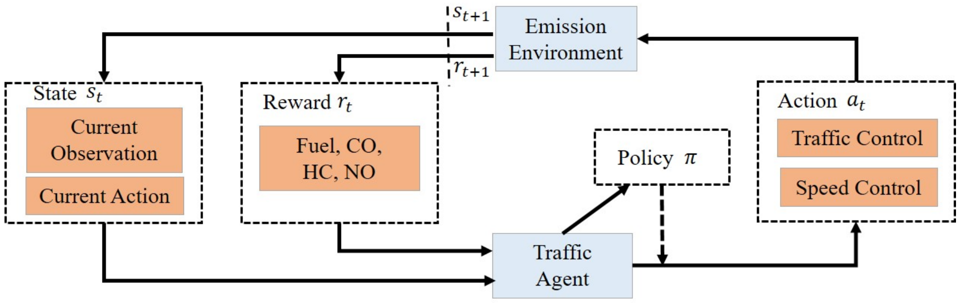

Fig. 2. Emission control policy problem definition

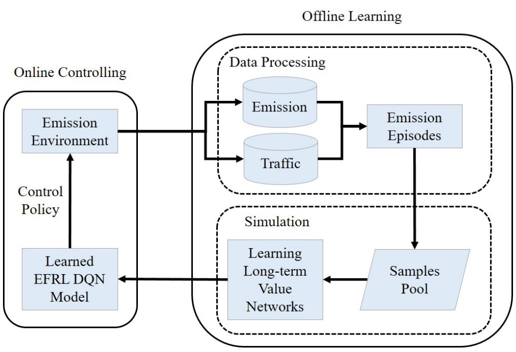

Fig. 3. The system framework of EFRL

$$\begin{aligned}\min_{\theta} L &= \mathbb{E}[(Q^*(s_t,a_t,\theta) - Q^*_{target}(s_t,a_t,\theta^-))^2] \\ &= \mathbb{E}[(Q^*(s_t,a_t,\theta) - (r_t + \gamma \cdot \max_{a_{t+1}} Q^*(s_{t+1},a_{t+1},\theta^-)))^2]\end{aligned} \tag{6}$$

And the pseudo algorithm for training the optimal long-term value network is detailed in Algorithm 1.

Algorithm 1 EFRL DQN training algorithm

Require:
Sample replay buffer D; emission episodes $EF = \{EF^1, EF^2, \cdots, EF^N\}$; episode length T; behavior network parameter θ; target network parameter θ^-.
Return: Learned EFRL DQN model $Q^*(\theta)$
$D \leftarrow \emptyset$//Initialize the replay buffer D
Random initialize $Q^*(S, A, \theta)$, $\theta^- \leftarrow \theta$
for each episode $EF^i (1 \leqslant i \leqslant N)$ **do**
  for each time step $t(1 \leqslant t \leqslant T)$ **do**
    random select an action a_t with probability $\epsilon \in [0,1]$// ϵ is a parameter for exploration
    otherwise select $a_t = arg\max_{a \in A} Q^*(s_t, a, \theta)$
    the traffic agent executes action a_t and transits to next state s_{t+1}, receiving reward r_t.
    storing new sample (s_t, a_t, s_{t+1}, r_t) to D
    sample random minibatch (s_t, a_t, s_{t+1}, r_t) from D
    $Q^*(s_t, a_t, \theta) = Q^*_{target}(s_t, a_t, \theta^-)$
    minimize L (Eq. 6)
    $\theta^- \leftarrow \theta$
  end for
end for

5. EXPERIMENTS

5.1 Data and setup

The proposed method is implemented in a high performance server with GeForce GTX 1080Ti GPU.

And we use real-world road segment to evaluate the emission control reinforcement learning model. Specifically, we collect emission data from vehicle remote sensing systems deployed on a main road in Hefei, which is authorised by Hefei Environmental Protection Bureau. As a sensor station may not have records sometimes, the missing entries are filled by average value. And the data details are summarized in Table 2.

5.2 Baselines and Metric

To evaluate the performance of the proposed EFRL model, we compare it with the following baselines.

Random policy (RP) This emission control policy means

Table 2. Experimental Emission Datasets

Dataset	Measurement
Location	Hefei
Time Span	2017/5/1-2017/7/31
Time interval	1 hour
Evaluation stations	4
Pollutants measurements	Fuel, CO, HC, NO
Traffic volume	[0,3512]$(V \cdot h^{-1})$
Vehicle speed	[0,77.9]$(km \cdot h^{-1})$
Available time interval	2091 (117 missing)

that the traffic agent selects a random action for traffic volume and traffic average speed limits policy each time.
Monte Carlo (MC) MC estimates optimal actions based on experience episodes without complete knowledge of the environment Sutton and Barto (2018). With this strategy, the traffic agent revises its policy and value estimates based on emission episodes experience.
Q-learning (QL) The QL is the most commonly used reinforcement learning algorithms Dayan and Watkins (1992)Mahadevan (1996). The QL agent chooses the optimal action based on the largest Q-value.

Metric evaluation To evaluate the EFRL control policy, we compare the total emission in average episode with the baseline policies. The average episode emission is defined as follows.

$$EF_{k,average} = \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{N}EF_k(i,t) \qquad (7)$$

where k$\in$ $\{Fuel, CO, HC, NO\}$, N is the total emission episodes, T is the episode length.

5.3 Results

In the emission policy evaluation experiment, we used the policy algorithms as described in previous section to run different traffic emission simulations with traffic and speed control. Fig.4 show the total emission of different emission controlling policies. And the emission reduction ratio respect to the total emission without controlling policy is shown in Table3, the emission reduction is little while taking the random policy. Moreover, the policies perform unstably on different pollution, i.e. MC and QL performs well on CO, HC reduction, MC and QL achieve similar results on NO reduction. And our EFRL achieves the best results on all pollution reduction.

Table 3. Emission reduction comparison on average episode

Control Policy	Fuel	CO	HC	NO
RP	↑ 1.28%	↑ 0.11%	↓ 5.54%	↓ 6.01%
MC	↓ 56.42%	↓ 76.86%	↓ 86.95%	↓ 61.93%
QL	↓ 59.34%	↓ 78.08%	↓ 96.38%	↓ 62.37%
EFRL	↓ 62.39%	↓ 89.54%	↓ 97.72%	↓ 65.56%

The traffic statistics of different policies are shown in Table 4 and Fig.5, comparing with the none control policy and random policy, the reinforcement learning based approaches are more stable on the traffic volume and

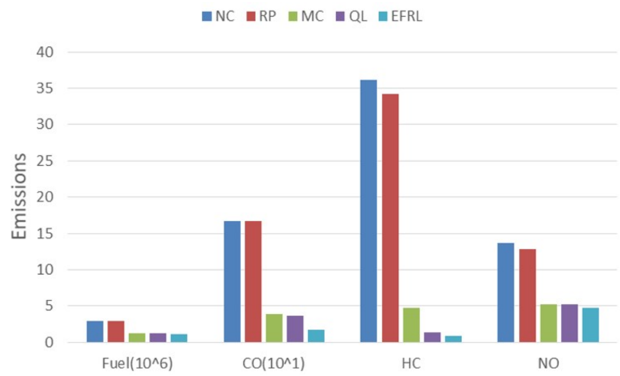

Fig. 4. Emission controlling comparison

Table 4. Traffic statistics of policies on Fuel

Models	Traffic volume		Speed	
	Mean	Std	Mean	Std
None control	569.02	278.29	11.55	1.87
RP	576.33	276.50	11.53	1.85
MC	316.89	122.39	11.58	0.53
QL	303.40	89.09	11.57	0.46
EFRL	266.85	81.99	11.15	1.46

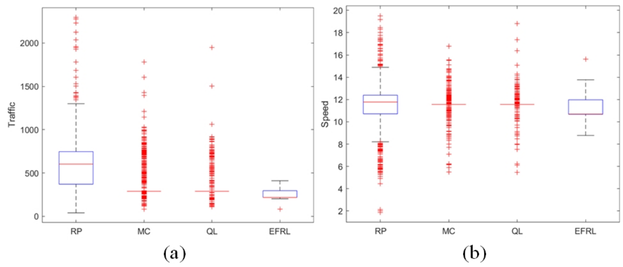

Fig. 5. Policy quality comparison on Fuel:(a) traffic volume distribution;(b) traffic speed distribution.

speed control. We can find that the EFRL policy can significantly reduce the traffic emission in average episode while ensuring a quite normal traffic volume and speed.

And to evaluate the training performance of EFRL, Fig.6 shows the total reward of training iterations. We can find that as the training process, there is a increasing trend of total reward.

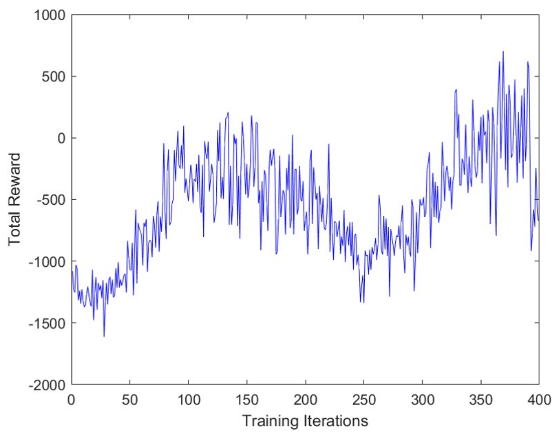

Fig. 6. Total reward of training iterations

6. CONCLUSION

In this paper, we solved the traffic emission management based on deep reinforcement learning model, namely EFRL. Different from the existing works which leverage the macroscopic traffic flow model and microscopic traffic emission model to simulate the emission environment, the EFRL directly learned from the history emission sequence, and used a compound emission state to capture the system temporal dependencies. To deal with the large state and action space, a DQN is applied to estimate the optimal long-term value function. The proposed policy is evaluated on real world vehicle emission data in Hefei. And the comparing result demonstrated the effectiveness of the proposed method.

In the future, we will extend the single road segment control policy to other spatiotemporal emission controlling on a regional scale.

REFERENCES

Bel, G., Bolanc, C., Guilln, M., and Rosell, J. (2015). The environmental effects of changing speed limits: A quantile regression approach. *Transportation Research Part D*, 36, 76–85.

Carslaw, D.C., Goodman, P.S., Lai, F.C.H., and Carsten, O.M.J. (2010). Comprehensive analysis of the carbon impacts of vehicle intelligent speed control. *Atmospheric Environment*, 44(23), 2674–2680.

Chao, L., Chen, H., and Grant-Muller, S. (2013). An indirect reinforcement learning approach for ramp control under incident-induced congestion. In *International IEEE Conference on Intelligent Transportation Systems*.

Cruciol, L.L.B.V., De Arruda, A.C., Weigang, L., Li, L., and Crespo, A.M.F. (2013). Reward functions for learning to control in air traffic flow management. *Transportation Research Part C*, 35(10), 141–155.

Dayan, P. and Watkins, C. (1992). Q-learning. *Machine learning*, 8(3), 279–292.

De Blasiis, M., Di Prete, M., Guattari, C., Veraldi, V., Chiatti, G., and Palmieri, F. (2014). The effects of traffic flow conditions on the pollutants emissions: A driving simulator study. *Advances in Transportation Studies*, 2(SPECIAL ISSUE), 59–70. doi:10.4399/97888548735377.

Dijkema, M., Zee, S.C.V.D., Brunekreef, B., and Strien, R.T.V. (2008). Air quality effects of an urban highway speed limit reduction. *Atmospheric Environment*, 42(40), 9098–9105.

Duell, M., Levin, M., and Waller, S. (2014). Urban vehicle energy consumption for policy evaluation: Impact of electric vehicles. 271–278.

Li, L., Lv, Y., and Wang, F.Y. (2016). Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica*, 3(3), 247–254.

Li, Y., Ukkusuri, S.V., and Fan, J. (2018). *Computers & Operations Research*, 99, 90–108.

Li, Z., Liu, P., Xu, C., Duan, H., and Wang, W. (2017). Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks. *IEEE transactions on intelligent transportation systems*, 18(11), 3204–3217.

Mahadevan, S. (1996). Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Machine learning*, 22(1-3), 159–195.

Miles, A., Zaslavsky, A., and Browne, C. (2018). Iot-based decision support system for monitoring and mitigating atmospheric pollution in smart cities. *Journal of Decision Systems*, 1–12.

Nishi, T., Otaki, K., Hayakawa, K., and Yoshimura, T. (2018). Traffic signal control based on reinforcement learning with graph convolutional neural nets. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 877–883. IEEE.

Panis, L.I., Beckx, C., Broekx, S., Vlieger, I.D., Schrooten, L., Degraeuwe, B., and Pelkmans, L. (2011). Pm, nox and co emission reductions from speed management policies in europe. *Transport Policy*, 18(1), 32–37.

Panis, L.I., Broekx, S., and Liu, R. (2006). Modelling instantaneous traffic emission and the influence of traffic speed limits. *Science of the Total Environment*, 371(1C3), 270–285.

Pérez, P., Trier, A., and Reyes, J. (2000). Prediction of pm 2.5 concentrations several hours in advance using neural networks in santiago, chile. *Atmospheric Environment*, 34(8), 1189–1196.

Shu, L., Schutter, B.D., Zegeye, S.K., Hellendoorn, H., and Xi, Y. (2013). Integrated urban traffic control for the reduction of travel delays and emissions. *IEEE Transactions on Intelligent Transportation Systems*, 14(4), 1609–1619.

Shuai, L., Hellendoorn, H., and Schutter, B.D. (2017). Model predictive control for freeway networks based on multi-class traffic flow and emission models. *IEEE Transactions on Intelligent Transportation Systems*, 18(2), 306–320.

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

Sutton, R.S., Barto, A.G., et al. (1998). *Introduction to reinforcement learning*, volume 2. MIT press Cambridge.

Uzunova, M., Losero, R., Lauber, J., and Djemai, M. (2012). Traffic velocity control for evaluation the impact of gases emissions: Case study of toll plaza. In *International Symposium on Environment Friendly Energies & Applications*.

Walraven, E., Spaan, M.T., and Bakker, B. (2016). Traffic flow optimization: A reinforcement learning approach. *Engineering Applications of Artificial Intelligence*, 52, 203–212.

Xu, Z., Cao, Y., and Kang, Y. (2019). Deep spatiotemporal residual early-late fusion network for city region vehicle emission pollution prediction. *Neurocomputing*, 355, 183–199.

Xu, Z., Kang, Y., and Cao, Y. (2020). Emission stations location selection based on conditional measurement gan data. *Neurocomputing*, 388, 170–180.

Yau, K.L.A., Qadir, J., Khoo, H.L., Ling, M.H., and Komisarczuk, P. (2017). A survey on reinforcement learning models and algorithms for traffic signal control. *Acm Computing Surveys*, 50(3), 1–38.

Zhang, L.D. and Zhu, W.X. (2015). Delay-feedback control strategy for reducing co2 emission of traffic flow system. *Physica A: Statistical Mechanics and its Applications*, 428, 481–492.