Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Urea Injection Control Based on Deep-Q Networks for SCR Aftertreatment Systems

Shin Young Bae * **Dong Hwi Jeong** * **Yeonsoo Kim** *
Byung Jun Lee * **Sanha Lim** * **Changho Jung** **
Chang Hwan Kim ** **Yong Wha Kim** ** **Jong Min Lee**† *

* *School of Chemical and Biological Engineering, Institute of Chemical Processes, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea (e-mail: jongmin@snu.ac.kr).*
** *Powertrain Performance Development Center, R&D Division, Hyundai Motor Company, 150, HyundaiYeonguso-ro, Hwaseong-si, Gyeonggi-do, 445-706, Republic of Korea.*

Abstract: The regulations on NOx emissions from diesel vehicles have been stringent in recent years. Various techniques such as lean NOx trap (LNT) and selective catalytic reduction (SCR) have been developed to lessen the NOx emissions . The urea-based SCR method, which utilizes NH_3 as reducing agent to remove NOx, is widely used. Determining optimal amount of injected urea that keeps NOx at outlet below regulated NOx emission and also minimizes the amount of dosed urea is important. Model predictive control (MPC) is popularly used to determine the optimal amount of injected urea. However, applying MPC to real vehicle driving may be difficult because the on-line computation of MPC is too costly to be conducted in the engine control unit (ECU), the computation performance of which is significantly low at present. Therefore, reinforcement learning (RL) is considered as an alternative to on-line control method. In this paper, deep Q-networks (DQN), which is an off-policy RL with discrete action space and suitable to solve high dimensional problem, is applied to determine the amount of urea injection in the SCR system. The simulation of urea injection control with DQN has been conducted with respect to inlet NOx emissions of real driving data.

Keywords: Diesel vehicle, Selective catalytic reduction, Urea-based SCR system, Reinforcement learning, Model-free learning, Deep-Q networks.

1. INTRODUCTION

Diesel engines have been widely used in the vehicle due to their superior fuel efficiency and lower CO2 emission. However, the diesel engines have a challenging limitation that it generates NOx, which is known to be harmful to the environment and human health (Morawska et al., 2004). With global increasing concerns on the issues, the regulations on the NOx emission from diesel vehicles have been stringent all over the world. According to the recent emission regulations of European Union, the limited amount of NOx was reduced from 180 mg/km at Euro 5 to 80 mg/km at Euro 6 (Ko et al., 2017). Considering the incoming emission regulation of Euro 7 which is expected to be more harsh and challenging, efforts to eliminate NOx will be even greater (Puškár and Kopas, 2018).

To satisfy the tightened regulation, various technologies to reduce NOx emission from diesel vehicle. Among those technologies, urea-based SCR is a promising method due to its ability to reduce NOx efficiently (Gabrielsson, 2004). In the urea-based SCR system, the urea is injected into the hot gas upstream of the catalytic reactor. The urea is thermally decomposed to NH_3, which is adsorbed into the catalyst in the reactor. The adsorbed NH_3 is utilized as a reductant, converting NOx to N_2 (Koebel et al., 1996).

Control of urea injection is play a important role in attaining high NOx removal efficiency. In addition, depending on the control method of urea injection, the amount of NH_3 slip at the tailpipe and remaining amount of urea in a tank are determined. Thus, many researchers have focused on the development of efficient control methods for urea injection in the SCR system. In particular, model predictive control (MPC) is one of the most widely used control methods in the urea-based SCR system. Chiang et al. propose a new controller design with MPC that minimizes NOx emissions and NH3 slip during the transient driving cycle (Chiang et al., 2010). Kim et al. suggest a control method integrating backstepping control with MPC, which makes SCR system stable with respect to disturbances and determines a desired adsorbed NH_3 coverage fraction of catalyst (Kim et al., 2018). The studies demonstrate that NOx at outlet controlled by MPC satisfies the regulatory amount. In addition, it is suggested that the overall efficiency of MPC, which considers NOx, NH_3 slip, and injected urea, is superior to other control methods.

Though MPC has been widely used as a fairly good controller for the urea-based SCR system, MPC has a critical limitation that it requires excessive on-line computation. Application of MPC, which should solve a challenging optimization problem on-line, might be difficult when the

mathematical model of the system is too complex or sampling time is short (Lee, 2014). The other disadvantage of MPC is the model-plant mismatch. In utilizing MPC, the plant is usually reduced to a simplified model where the obtained policy is not equivalent to the optimal policy of the real system (Badwe et al., 2009).

The urea-based SCR system is a complicated system where dynamics inside the SCR reactor are represented as a partial differential equations (PDE) with respect to bulk gas and temperature. The dimension of states in the system can be fairly high in dealing with the infinite states of PDE. Thus, the model complexity in the urea-based SCR system poses a challenge of on-line computation in MPC. To lessen the limitations of the on-line computation, MPC usually utilizes reduced models such as control oriented two-cell model. (Kim et al., 2018).

Nevertheless, from a practical point of view, the feasibility of computation of MPC in real driving is still doubtful. The main reason is that engine control unit (ECU), where the computation of control performs in the modern vehicles, still lacks ability to conduct complex calculation. The demanding on-line computation, for example, from MPC is difficult to implement due to the insufficient number of ECUs, the footprint of which is limited in the modern vehicle (Stewart and Borrelli, 2008). It is also suggested that the on-line computation of MPC would be impractical due to poor performance of ECU (Del Re et al., 2010). In addition, the sampling interval for measurement of incoming NOx is short, which makes application of MPC to real driving more difficult.

Reinforcement learning (RL) (Sutton and Barto, 2018), an off-line approach, can be an alternative to MPC for urea based SCR system. The advantage of RL as a controller is that computation time is fairly short because it utilizes a control policy obtained off-line, not in each time step (Lee, 2014). Among numerous RL algorithms, this study employs the deep-Q networks (DQN), an off-policy RL with discrete action space (Mnih et al., 2015) because DQN is a proper algorithm in dealing with high dimensional system such as urea based SCR system. It is also expected that the mismatch problem can be alleviated in that DQN is model-free based method.

The rest of the paper is organized as follows: In Section 2, overall mathematical models of SCR system are presented. In Section 3, detailed explanation on DQN algorithm is provided. In Section 4, simulation results on the SCR system are presented to validate the performance of DQN. The results of cost and NOx emissions at outlet with DQN and MPC will be compared. The computation time of DQN and MPC is also compared to discuss performance of DQN controller.

2. PRELIMINARIES

2.1 Virtual plant of SCR system

The SCR system consists of honeycomb catalytic monolith. The pollutant bulk gas containing NOx passes through each channel of monolith, where the NOx is removed by reduction with the adsorbed NH_3 in the catalyst. The inside of each channel is covered with the porous layers containing catalyst, called washcoat.
The governing equations of the virtual SCR system model is referred to (Kim et al., 2018). In addition, the following assumptions are required to establish governing equations:

- Plug flow and negligible pressure drop
- Incompressible gas
- Negligible heat transfer between the SCR system and ambient system
- The bulk gas temperature and wash coat temperature are the same

The governing equations can be written as

- Concentration of bulk gas species ($x_{g,j}$, m=catalyst)

$$\phi_g\left(\frac{\partial x_{g,j}}{\partial t} + u\frac{\partial x_{g,j}}{\partial z}\right) = -k_{m,j}G_a\left(x_{g,j} - x_{wc,j}\frac{T_g}{T_m}\right). \quad (1)$$

- Concentration of gas species in washcoat ($x_{wc,j}$)

$$(1-\phi_g)\phi_{wc}\epsilon_{wc}\frac{\partial x_{wc,j}}{\partial t} = k_{m,j}G_a\left(x_{g,j}\frac{T_m}{T_g} - x_{wc,j}\right) + \frac{RT_m}{P}\sum_{i=1}^{n}\lambda_{ji}r_i. \quad (2)$$

- Coverage fraction of catalyst ($\theta_{m,k}$, k=site)

$$\frac{\partial \theta_{m,k}}{\partial t} = \frac{\sum_{i=1}^{n}\lambda_{ki}r_i}{\psi_k} \quad (3)$$

- Temperature of bulk gas (T_g)

$$\phi_g\rho_g c_{p,g}\left(\frac{\partial T_g}{\partial t} + u\frac{\partial T_g}{\partial z}\right) = -k_h G_a(T_g - T_m) \quad (4)$$

- Temperature of gas in washcoat (T_m)

$$(1-\phi_g)(1-\phi_{wc}\epsilon_{wc})\rho_m c_{p,m}\frac{\partial T_m}{\partial t} = k_h G_a(T_g - T_m) - \sum_{i=1}^{n}H_j^f\lambda_{ji}r_i \quad (5)$$

where i, j, and k are the reaction index, species index, and catalyst site index, respectively. ϕ_g and ϕ_{wc} are the volume ratio of gas layer and washcoat layer to the reactor, u denotes the bulk gas velocity, $k_{m,j}$ is the mass transfer coefficient of the each species, G_a is a ratio of wetted area to the reactor volume, ϵ_{wc} denotes the washcoat porosity, λ_{ji} is the reaction coefficient, r_i is the reaction rate of the ith reaction, ψ_k is the storage capacity of k site, and H_j^f denotes the enthalpy of formation of each species.

2.2 Control oriented two-cell model

In this paper, MPC is utilized to initialize the policy in the training. As discussed above, MPC usually uses the reduced model from the plant to lessen the computation load. The control oriented two-cell model is a popular simplified model for MPC. The control oriented model assumes that SCR system is the continuous stirred tank reactor (CSTR). Assuming all the states in the SCR systems are homogeneous and neglecting the energy balance, the model can be represented as a set of 0-D ordinary differential equations (ODE). It is reasonable to split the

Table 1. Reactions and Kinetics of SCR system

Reaction	Kinetics
$NH3 + S1 \leftrightarrow NH3 - S1$	$r_1 = k1_{1f} x_{NH_3}(1-\theta)\psi - k_{1b}\theta\psi$
$2NH_3 - S1 + \frac{3}{2}O_2 \rightarrow N_2 + 3H_2O + 2S1$	$r_2 = k_2 x_{O_2}\theta\psi$
$NO + \frac{1}{2}O_2 \leftrightarrow NO_2$	$r_3 = k_{3f} x^0_{O_2}.5x_{NO} - k_{3b}x_{NO_2}$
$4NH_3 - S1 + 4NO + O_2 \rightarrow 4N_2 + 6H_2O + 4S1$	$r_4 = k_4 x_{NO}\theta\psi$
$2NH_3 - S1 + NO + NO_2 \rightarrow 2N_2 + 3H_2O + 2S1$	$r_5 = k_5 x_{NO} x_{NO2}\theta\psi$
$4NH_3 - S1 + 3NO_2 \rightarrow 3.5N_2 + 6H_2O + 4S1$	$r_6 = k_6 x_{NO_2}\theta\psi$
$2NH_3 - S1 + 2NO_2 \rightarrow N_2 + N_2O + 3H_2O + 2S1$	$r_7 = k_7 x_{NO_2}\theta\psi$

control oriented model into several cells because the model with only a single section is not accurate to represent all the length of the SCR reactor. Among the proposed SCR cell models (Schär et al., 2004), (Upadhyay and Van Nieuwstadt, 2002), we adopted the two cell model, which is utilized for controlling urea injection (Kim et al., 2018). The equations of the two cell model are represented as follows:

- Concentration of bulk gas species($x_{g,j}$)

$$\frac{\mathrm{d}x_{g,j}}{\mathrm{d}t} = \frac{1-\phi_g}{\phi_g}\sum_{i=1}^{n}\lambda_{ji}r_i - u\frac{x_{g,j} - x_{g,j,inlet}}{L} \tag{6}$$

- Coverage fraction of catalyst($\theta_{m,k}$, k=site)

$$\frac{\mathrm{d}\theta_m}{\mathrm{d}t} = \frac{\sum_{i=1}^{n}\lambda_i r_i}{\psi} \tag{7}$$

where j is each species of the bulk gas, NO, NO_2, O_2, and NH_3 and L is the length of SCR system.

2.3 Reactions and kinetics of SCR system

The reactions and kinetics of the SCR system are adopted from (Olsson et al., 2008) and shown in Table 1.

3. DEEP Q-NETWORKS

RL aims to find a nearly optimal control policy by interacting with environment and learning the optimal value function and/or corresponding policy with reinforcement signal such as cost or reward. For state x_t in each time step t, the agent chooses an action u_t with given policy π, which maps state x_t to action u_t. The state is transitioned with state dynamics $x_{t+1} = f(x_t, u_t)$ and the agents receives a reward $r_t(x_t, u_t)$. The return is often defined as the sum of discounted future rewards $G(t) = \sum_{i=t}^{\infty}\gamma^{i-t}r_i(x_i, u_i)$, where γ is a discount factor. During the training, the agent selects an action to maximize the expectation of returns for each state x_t.
Q-value is defined as an expectation of returns at state x_t with action u_t and is represented as

$$Q^\pi(x_t, u_t) = \mathbb{E}[G(t)|x_t, u_t] \tag{8}$$

where $Q^\pi(x_t, u_t)$ denotes the Q-value at a state x_t and an action u_t, which are derived from the policy π.
By applying Bellman equation, the Equation (8) also can be represented in a recursive relationship as follows:

$$Q^\pi(x_t, u_t) = \mathbb{E}[r_t + \gamma\mathbb{E}[Q^\pi(x_{t+1}, u_{t+1})]] \tag{9}$$

Q-learning utilizes Equation (9) to iteratively update the Q-value by applying the off-policy method represented as

$$\begin{aligned}(Q^\pi)'(x_t, u_t)) =& Q^\pi(x_t, u_t) + \alpha(r_t + \gamma\max_{u_{t+1}}Q^\pi(x_{t+1}, u_{t+1}) \\ &- Q^\pi(x_t, u_t))\end{aligned} \tag{10}$$

where $0 < \alpha < 1$ is the learning-rate parameter and $(Q^\pi)'(x_t, u_t)$ denotes the updated Q-value (Watkins and Dayan, 1992).
Model-free RL algorithms, including Q-learning, is an approach that learns the policy only from sequences of samples obtained during episodes. In other words, it does not need any information about the system dynamics in determining the action u_t (Gu et al., 2016). There are two advantages of the model-free approach over the model-based approach. First, the model-free approach needs less amount of on-line computation load because it follows the policy, which is presented in a tabular or parameterized form. (Lee, 2014). Second, it can avoid model error, which occurs in linearization or reduction of the model in the model-based approach (Badwe et al., 2009). Thus, model-free approach might be appropriate for SCR system where on-line computation load is demanding with the limited capacity of ECU and the system is too complex, which is required to be simplified.
In most of the problems which we would like to utilize RL, the state space is high dimensional. Accordingly, the Q-learning based on a tabular search is impossible to find an optimal action due to the the curse-of-dimensionality. Thus, it is required to approximate Q-value with given data of states and actions in the parametric form. (**?**).
Deep Q-networks (DQN) approximates the Q-value by applying deep neural networks. The notable technique of DQN is the usage of experience replay (U) where the sequences of samples (x_t, u_t, r_t, x_{t+1}) at each time are stored. The samples randomly chosen from experience replay are used to update Q-value, which removes correlations between consecutive samples and hinders the parameters of Q-value to be stuck in a poor local minimum (Mnih et al., 2013). The loss function for Q-learning updates at iteration i is

$$\begin{aligned}L_i(\theta_i) = \mathbb{E}_{(x_t, u_t, r_t, x_{t+1})\sim U}[(r_t + \gamma\max_{u_{t+1}}Q^\pi(x_{t+1}, u_{t+1}; \theta_i^-) \\ - Q^\pi(x_t, u_t; \theta_1))^2]\end{aligned} \tag{11}$$

where θ_i are the parameters of the Q-network at iteration i and θ_i^- are the parameters of the target Q-network at iteration i, which are only replaced by the parameters of

the Q-network θ_i at a specific iteration step.
In several studies, it has been demonstrated that DQN learns a nearly-optimal policy successfully. Atari 2600 games was tested by DQN, which is trained with high-dimensional input pixels, whose output is the estimation of Q-value. The result showed that the performance of DQN exceeded those of previous algorithms, also being a level matching the professional human testers in some games (Mnih et al., 2015). DQN was also applied to a controller of robot manipulator with no prior knowledge, whose objective is target reaching. The experiment proved that the controller, which is trained by DQN, performed the target reaching successfully (Zhang et al., 2015).

4. SIMULATION RESULTS AND DISCUSSION

To validate the performance of DQN, simulation results on the SCR system with time-varying inlet emissions are presented. The time-varying inlet emissions data were obtained by experiments of FTP75 cycle, which is widely used for the real driving mode (Zhang et al., 2014).
In this simulation, virtual plant of SCR system mentioned in Section 2.1 was used and the inlet gas was considered as a disturbance variable. We assumed the complete thermolysis of urea which is injected into the hot upstream of SCR system. The state variables include concentrations of gas species, temperature of gas and coverage fraction of catalyst, and the state dimension is 48. It is assumed that all the states are measurable. The number of action is one, which is the amount of injected urea and the value of which is bounded with upper and lower limits.

4.1 Settings for deep Q-network

Actions in DQN model are assumed to be discrete. Because it is infeasible to choose an action by comparing Q-values of infinite actions, the number of actions was limited to 40 in this study.
The number of hidden layers for approximating Q-function are three and each layer consists of 50, 50, and 30 nodes, referring to the network structure of (Kim et al., 2020). Both states and an action are input arguments of the networks and Q-value is estimated as an output. As an activation function, Relu function was used, which guarantees a short training time than other activation functions (Krizhevsky et al., 2012). To prevent an over-fitting problem that often occurs in making deep neural networks model, methods of L2-regularization (Bilgic et al., 2014) are applied to establish our model. In updating parameters of the neural networks, Adam-optimizer (Kingma and Ba, 2014) was utilized.
The construction of cost function is significantly important in establishing DQN model (Russell, 1998). In the case of SCR system, reducing the amount of NOx emission at the outlet below the regulated quantity is the most important. Ammonia slip at the tailpipe, which is not utilized as a reductant, also should be avoided. In addition, minimizing the amount of injected urea is significant. Improperly overdosed urea leads to a rise of operating cost and inconvenience of urea make-up. Considering the objectives of SCR system and the facts that the incoming regulation on the emissions will be more stringent, a cost function can be formulated as

$$\begin{aligned} r(t) =& \|x_{NO}(t)\|^2_{Q_1} + \|x_{NO_2}(t)\|^2_{Q_2} + \|x_{NH_3}(t)\|^2_{Q_3} \\ &+ \|u_{NH_3}(t)\|^2_R \end{aligned} \tag{12}$$

where Q_1, Q_2, Q_3, and R is 5, 5, 2.5, and 0.5 respectively and the concentrations of gas species are scaled value. A relatively high values are weighted to the concentration of NOx to satisfy regulation on NOx emission preferentially.

4.2 Learning process

As discussed above, parameters of Q-function are updated by sequences of samples (x_t, u_t, r_t, x_{t+1}) induced by an action with given policy at each time. For the training of DQN, the number of episodes was set as 700. For each episode, the horizon is 1200 seconds and the sampling time is one second.
An exploration plays an important role in the training of model-free RL algorithm such as DQN. To find better actions which might give lower costs, it is necessary to try actions that have not been met before (Sutton and Barto, 2018). The exploration was also conducted in training DQN for SCR systems. In the earlier period of the training, exploratory noises was added to the actions derived from DQN. The size of the added noise was set to decrease as the training proceeds. We also examined the appropriate number of episodes, where enough exploration is available, with several trials. The exploration was employed during the first a hundred episodes. After 100th episodes, actions were chosen by the DQN policy without noises.

4.3 Results and analysis

Figure 1 describes the sum of costs, represented as $\sum_{t=1}^{1200} r(t)$, in each episode during the learning. The discount factor was not included in obtaining the sum of costs. During the first a hundred episodes, it is shown that the sum of costs remains high and fluctuates considerably. The reason is that DQN explored various actions, which might result in higher sum of costs. After the periods of exploration, the noises added to actions are eliminated, which make fluctuations in the sum of costs lessened. From the approximately 400th episode, it can be found that the sum of the costs from DQN are stable and remained below the value from MPC. This indicates that DQN converged to optimal policy and the performance of DQN is superior to MPC in terms of the given cost function. As it will be discussed later, it appears that small usage of injected urea is the main reason for reducing the sum of costs from DQN.
Figure 2 indicates the scaled amount of injected NH_3 determined by DQN and MPC. At this time, DQN model, whose training was finished after 700 episodes, was used. It can be seen that DQN yields a fairly different amount of urea with that of MPC. As with urea based SCR system, MPC usually utilizes a simplified model from a complex real system, leading to failure of obtaining optimal solution for real system (Badwe et al., 2009). On the other hand, DQN is a controller that is approximated with data from real system, not model. Thus, it is acceptable that the actions from DQN might consider a real system better than MPC. The results showed that the cumulative amount of

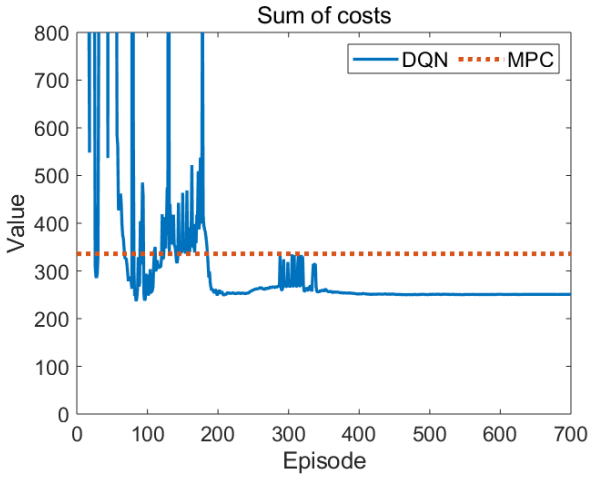

Fig. 1. Sum of costs for each episode in DQN and MPC.

injected NH_3 from DQN is 12 percent less than that of MPC.

Figure 3 shows the scaled cumulative amount of NOx at outlet controlled by DQN and MPC. The limited amount of NOx emissions by EURO 6 regulation is also depicted in Figure 3. The cumulative regulated NOx emissions were calculated with respect to the driving distance of FTP75 cycle. It is observed that NOx emission at outlet controlled by DQN is nearly comparable to that of MPC. Overall, the removal of NOx by DQN is satisfactory except the periods from 180 to 280 seconds. In this period, it is thought that considerable amount of NOx suddenly enters the SCR system while the coverage of NH_3 are small, where the performance to reduce the NOx is insufficient. Thus, the improvement on DQN controller to handle the sudden influx of large NOx emission is required.

Figure 4 depicts the scaled concentration of NH_3 slip at tailpipe controlled by DQN and MPC. It is shown that NH_3 slip by DQN is nearly comparable to that of MPC. For the overall periods, the amount of NH_3 slip controlled by DQN does not exceed the reference value which guarantees the satisfactory emission control (Hsieh and Wang, 2011). Thus, it seems that performance of DQN controller is acceptable in handling the NH_3 slip.

Finally, we compared a computation time between DQN and MPC. The simulation was conducted on an Intel Core i7-6700 3.40GHz processor, 32GB RAM, and a GeForce RTX-2070 graphic card. During 1200 seconds, the average computation time of DQN and MPC per sampling time is 0.0018 s and 0.053 s, respectively. DQN could determine the amount of injected urea 30 times faster than MPC does in online. As discussed above, the computation performance of ECU in the modern vehicles is fairly poor. Thus, it is expected that DQN can be a potential alternative to MPC in controlling amount of urea injection for the real driving.

5. CONCLUSION

In this study, we applied DQN to controlling the amount of injected urea in the urea based SCR system. Through the simulations of urea injection control with DQN, it is demonstrated that computation time of DQN is fairly shorter than that of MPC and the sum of costs of DQN is less than that of MPC. These results that DQN has a

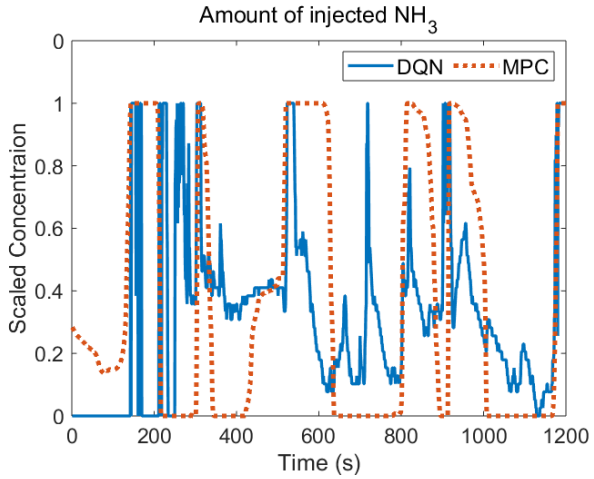

Fig. 2. Scaled concentration of injected NH_3 by DQN and MPC

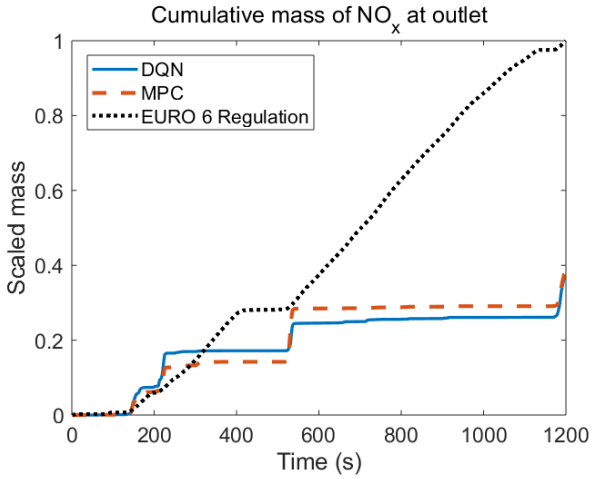

Fig. 3. Scaled cumulative amount of NO_x at outlet for DQN, MPC and EURO 6 Regulation

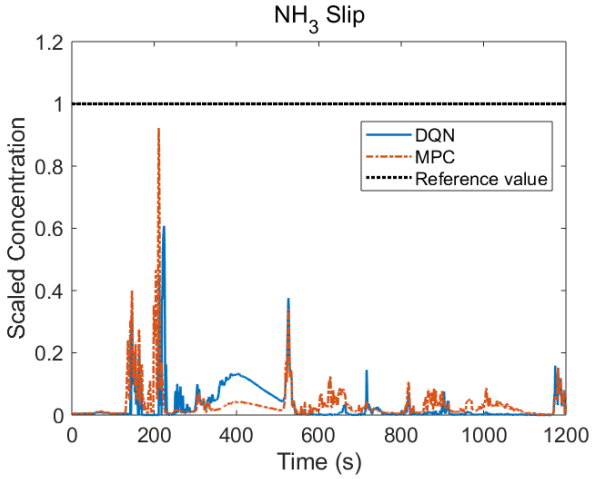

Fig. 4. Scaled concentration of NH_3 at outlet for DQN, MPC and reference value

possibility to control urea injection in the real driving, the performance of computation of which is poor at present. However, DQN controller failed to remove NOx efficiently when considerable amount of NOx suddenly injected to the SCR system. Thus, the coverage fraction which is thought to be important factor to prevent those situations will be considered in constructing DQN controller in the next study.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (MSIP) (NRF-2016R1A5A1009592).

REFERENCES

Badwe, A.S., Gudi, R.D., Patwardhan, R.S., Shah, S.L., and Patwardhan, S.C. (2009). Detection of model-plant mismatch in mpc applications. *Journal of Process Control*, 19(8), 1305–1313.

Bilgic, B., Chatnuntawech, I., Fan, A.P., Setsompop, K., Cauley, S.F., Wald, L.L., and Adalsteinsson, E. (2014). Fast image reconstruction with l2-regularization. *Journal of magnetic resonance imaging*, 40(1), 181–191.

Chiang, C.J., Kuo, C.L., Huang, C.C., and Lee, J.Y. (2010). Model predictive control of scr aftertreatment system. In *2010 5th IEEE Conference on Industrial Electronics and Applications*, 2058–2063. IEEE.

Del Re, L., Ortner, P., and Alberer, D. (2010). Chances and challenges in automotive predictive control. In *Automotive model predictive control*, 1–22. Springer.

Gabrielsson, P.L. (2004). Urea-scr in automotive applications. *Topics in catalysis*, 28(1-4), 177–184.

Gu, S., Lillicrap, T., Sutskever, I., and Levine, S. (2016). Continuous deep q-learning with model-based acceleration. In *International Conference on Machine Learning*, 2829–2838.

Hsieh, M.F. and Wang, J. (2011). A two-cell backstepping-based control strategy for diesel engine selective catalytic reduction systems. *IEEE Transactions on Control Systems Technology*, 19(6), 1504–1515.

Kim, J.W., Park, B.J., Yoo, H., Oh, T.H., Lee, J.H., and Lee, J.M. (2020). A model-based deep reinforcement learning method applied to finite-horizon optimal control of nonlinear control-affine system. *Journal of Process Control*, 87, 166–178.

Kim, Y., Jung, C., Kim, C.H., Kim, Y.W., and Lee, J.M. (2018). Backstepping control integrated with model predictive control for selective catalytic reduction system of diesel vehicle. In *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, 1376–1382. IEEE.

Kingma, D.P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Ko, J., Jin, D., Jang, W., Myung, C.L., Kwon, S., and Park, S. (2017). Comparative investigation of nox emission characteristics from a euro 6-compliant diesel passenger car over the nedc and wltc at various ambient temperatures. *Applied energy*, 187, 652–662.

Koebel, M., Elsener, M., and Marti, T. (1996). Nox-reduction in diesel exhaust gas with urea and selective catalytic reduction. *Combustion science and technology*, 121(1-6), 85–102.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105.

Lee, J.H. (2014). From robust model predictive control to stochastic optimal control and approximate dynamic programming: A perspective gained from a personal journey. *Computers & Chemical Engineering*, 70, 114–121.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529.

Morawska, L., Moore, M.R., and Ristovski, Z.D. (2004). Health impacts of ultrafine particles: Desktop literature review and analysis. *Report to the Australian Department of the Environment and Heritage*.

Olsson, L., Sjövall, H., and Blint, R.J. (2008). A kinetic model for ammonia selective catalytic reduction over cu-zsm-5. *Applied Catalysis B: Environmental*, 81(3-4), 203–217.

Puškár, M. and Kopas, M. (2018). System based on thermal control of the hcci technology developed for reduction of the vehicle nox emissions in order to fulfil the future standard euro 7. *Science of the Total Environment*, 643, 674–680.

Russell, S.J. (1998). Learning agents for uncertain environments. In *COLT*, volume 98, 101–103.

Schär, C.M., Onder, C.H., Geering, H., and Elsener, M. (2004). Control-oriented model of an scr catalytic converter system. Technical report, SAE Technical Paper.

Stewart, G. and Borrelli, F. (2008). A model predictive control framework for industrial turbodiesel engine control. In *2008 47th IEEE Conference on Decision and Control*, 5704–5711. IEEE.

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

Upadhyay, D. and Van Nieuwstadt, M. (2002). Modeling of a urea scr catalyst with automotive applications. In *ASME 2002 international mechanical engineering congress and exposition*, 707–713. American Society of Mechanical Engineers Digital Collection.

Watkins, C.J. and Dayan, P. (1992). Q-learning. *Machine learning*, 8(3-4), 279–292.

Zhang, F., Leitner, J., Milford, M., Upcroft, B., and Corke, P. (2015). Towards vision-based deep reinforcement learning for robotic motion control. *arXiv preprint arXiv:1511.03791*.

Zhang, H., Wang, J., and Wang, Y.Y. (2014). Application of nmpc on optimization of ammonia coverage ratio references in two-can diesel scr systems. In *2014 American Control Conference*, 220–225. IEEE.