CHEMICAL ENGINEERING TRANSACTIONS
VOL. 72, 2019
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Jeng Shiun Lim, Azizul Azri Mustaffa, Nur Nabila Abdul Hamid, Jiří Jaromír Klemeš

ISBN 978-88-95608-69-3; **ISSN** 2283-9216
DOI: 10.3303/CET1972016

The Real-Time Traffic Signal Control System for the Minimum Emission using Reinforcement Learning in Vehicle-to-Everything (V2X) Environment

Jooyoung Kim[a], Sangchul Jung[b], Kwangsik Kim[c], Seungjae Lee [b,*]

[a]Institute of Urban Sciences, University of Seoul, Korea
[b]Department of Transportation Engineering, University of Seoul, Korea
[c]Department of Public Administration, Sungkyunkwan University, Korea
sjlee@uos.ac.kr

As the population and vehicle ownership increase, emission of pollutants is also increasing. The percentage of GHG emission by transportation sector is about 21 % in 2015 (OECD), and this may be caused by frequent stop-and-go phenomenon or delay time of vehicles in signalized intersection. Generally, these could be minimised by driving in constant speed or decreasing the delay times with an efficient traffic signal control. On the other hand, researches try to decrease vehicles' delay time and to exclude the unnecessary stop-and-go phenomenon in an urban signalized intersection with an advent of V2X (Vehicle-to-Everything) technology development. Especially, in traditional pre-timed traffic signal control situation, even the autonomous vehicles would be impossible to exhibit their own maximum performance. Thus, the development of the traffic signal control system could have effects not only on the traffic flow but also on environmental aspects, which optimizes the signalized traffic flow based on the real-time vehicle information. In this research, on the premise of V2X environment, changes in traffic flow and the emission are analysed based on microscopic traffic information. In specific, the reinforcement learning model is constructed based on Deep Learning which learns the real-time traffic information and displays the optimal traffic signal. The performance of the system was analysed through microscopic traffic simulator - Vissim. The proposed system is expected to contribute on analysing the traffic flow and the environmental effects. Also, it is expected to contribute on constructing the green smart cities with an advent of autonomous vehicle operation in future V2X environment.

1. Introduction

Traffic flow is classified to continuous flow and signalized flow. The continuous flow, such as highway, is affected mainly by surrounding vehicles or geometric road design, while the signalized flow is controlled by the traffic signal control system. While there are delays arisen by the intervehicle interference in the continuous flow, they can be decreased by maintaining the optimal headway or detecting the unexpected situation through V2X technology which communicates between vehicles and infrastructure and controlling the vehicle. Delays in the signalized flow are mainly consist of stop delay, acceleration/deceleration delay and stop-and-go phenomenon which caused by the traffic signal control system, and these lead to increase vehicle emission. A careful traffic signal control is required to decrease delay and vehicle emission. Generally, pollutants on a road environment are produced by driving, acceleration/deceleration and stop of vehicle. Driving conditions of vehicle and their corresponding environmental influences are described in Table 1. Acceleration/deceleration and stop of vehicle conceptually comprise the delay time of vehicle and it can be said that the traffic signal control in an intersection decides an amount of pollutants. Therefore, the traffic signal controller which minimises the delay time and the emission of vehicle is required. The traditional traffic signal control methods are pre-timed signal control, actuated signal control and semi-actuated signal control. Descriptions of traditional traffic signal control methods are presented in Table 2.

Paper Received: 15 March 2018; Revised: 09 September 2018; Accepted: 06 December 2018

Please cite this article as: Kim J., Jung S., Kim K., Lee S., 2019, The real-time traffic signal control system for the minimum emission using reinforcement learning in v2x environment, Chemical Engineering Transactions, 72, 91-96 DOI:10.3303/CET1972016

It is hard for traditional traffic signal controllers to respond to real-time traffic condition because they depend on pre-processed data of traffic volume, pedestrian volume, or geometric design of road to decide phase plan.

Table 1: The environmental influence corresponding to the driving condition of vehicle

Driving condition	Environmental influence
Driving	An amount of pollutant emission is different by vehicle speed
Acceleration	Always consumes more fuel than driving and emission increases
Deceleration	A mechanical deceleration increases the fuel consumption
Stop	Pollutants increase as the stopping time gets longer

Table 2: Traditional traffic signal control methods

Control method	Description	Feature	Application
Pre-timed	Controls according to the pre-decided phase plan	Phase plan is decided according to the traffic and pedestrian volume	An area where similar traffic pattern repeats
Actuated	All lanes of all approaches are monitored by detectors and controlled	Phase sequence, green allocations and cycle length changes	An area where the traffic pattern fluctuates
Semi-actuated	Minor road or left turn movement of major road is detected and controlled	Green is on the major road unless a 'call' on the minor street is noted	A side street intersects with major arterial

It may result in ineffective traffic signal control and increase the emission. On the other hand, the real-time traffic signal control methods calculate the optimal signal plan based on the information collected by a road-side unit or V2X technology. Especially, the signal timing optimization through machine learning is possible to be applied to various traffic conditions and is possible to optimize various objective functions.

Researches which tried to decrease the vehicle delay and emission through traffic signal control is as follows: Li and Shimamoto (2011) proposed the method which decreases CO_2 emission and vehicle delay time through decision tree based on the vehicle information collected in the single intersection. As a result, CO_2 emission decreased 26.9 % compared to pre-timed signal controller. However, it could be impossible to apply to more complicated signal phase because of just considering the signalized intersection consisted of single lane and straight movement. Chitradevi et al. (2013) developed the controller based on branch-and-bound algorithm and compared to the adaptive fuzzy controller. The result showed decreased delay time and CO_2 emission. Jinpeng et al. (2013) used MOVES (Motor Vehicle Emission Simulation), which is pollutant emission estimation model based on second-by-second information of vehicle and proposed genetic algorithm to minimise delay time and pollutant emission. The trade-off relationship between delay time and emission rate of vehicle was also showed. Park et al. (2008) pointed out limitations of the existing macroscopic emission model (estimation based on the average speed of vehicles) and raised the necessity of using microscopic vehicle information. Pol (2016) developed a traffic signal controller for independent and coordinated intersections using Deep reinforcement learning. However, the intersection is simply composed of a single straight movement for all directions, and the state of the intersection is represented by a binary position matrix indicating the position of the vehicle as 0 or 1. Therefore, when the intersection becomes complex, practical application could be difficult.

Based on implications drawn from literature review, some points that should be satisfied are as follows:

- Setting the microscopic emission estimation function based on the individual vehicle information
- Establishing the objective function which minimises the delay time and emission simultaneously
- Determining the phase plan in real-time through optimizing the objective function
- Developing the traffic signal control system which can handle all movements in an intersection

In this paper, the proposed system that expresses the optimal signal uses each vehicle's information collected in V2X environment. This information is collected in real-time and is used to process each vehicle's delay and an amount of emission. As will be stated in section 2, estimating an amount of vehicle's emission is performed for each individual vehicle in the road network in real-time, not as using the vehicle's driving profile during analysis period. This could provide more accurate data and could make it possible to control the traffic signal in real-time. The reinforcement learning, which is the core method for controlling the traffic signal, uses this information to obtain the optimal signal for both minimum vehicle delay and emission rate. As Mnih et al. (2015) shows, the reinforcement learning is model-free method, which makes it possible to apply the learned model to various environment. It is expected that the proposed model has both accuracy and versatility.

2. Methodology

2.1 Vehicle emission estimation function

For estimating each vehicle's emission based on each vehicle's information, there exists MOVES. However, MOVES requires friction factors of each vehicle's driving condition, thus there exists difficulty in achieving data. In addition, MOVES analyses macroscopic range based on vehicle's driving profile during analysis period. Therefore, it is judged that using MOVES to develop real-time traffic signal controller is inappropriate. On the other hand, Ahn et al. (2002) introduced the method for estimating fuel consumption and emission based on vehicle's instantaneous speed and acceleration. This method is possible to be applied to solve real-time traffic signal control problem, because each vehicle's information can be achieved through the microscopic simulator. Estimation function of HC emission based on the vehicle's instantaneous speed and acceleration/deceleration is as Eq(1).

$$e = exp\left(\sum_{y=0}^{3}\sum_{x=0}^{3} k_{xy} * s^x * a^y\right) \quad (1)$$

In Eq(1), e is the emission rate of HC in mg/s, k is the model regression coefficients, s is the velocity in m/s and *a* is the acceleration in m/s^2. Details for model regression coefficients are shown in Table 3.

Table 3: Model regression coefficients

k_{xy}		x			
		x^0	x^1	x^2	x^3
Positive acceleration y	y^0	-0.876050	0.036270	-0.00045	2.55 x 10^{-6}
	y^1	0.081221	0.009246	-0.00046	4.00 x 10^{-6}
	y^2	0.037039	-0.006180	2.96E-04	-1.86 x 10^{-6}
	y^3	-0.002550	0.000468	-1.79 x 10^{-5}	3.86 x 10^{-8}
Negative acceleration (Deceleration) y	y^0	-0.755840	0.021283	-0.00013	7.39 x 10^{-7}
	y^1	-0.009210	0.011364	-0.0002	8.45 x 10^{-7}
	y^2	0.036223	0.000226	4.03 x 10^{-8}	-3.5 x 10^{-8}
	y^3	0.003968	-9 x 10^{-5}	2.42 x 10^{-6}	-1.6 x 10^{-8}

When the vehicle approaches to intersection with decelerating and accelerates for leaving the intersection, a large quantity of emission is produced. When the vehicle stops at the stop line, an instantaneous speed and acceleration is 0, thus a small amount of emission produced. Therefore, it can be said that there is a trade-off relationship between delay time and emission, as Jinpeng et al. (2013) showed. Because minimizing only the emission could rather increase the delay time, the objective function which minimises the delay time and emission is required.

2.2 The reinforcement learning model

The reinforcement learning model for optimizing the phase plan based on the individual vehicle information was constructed. The reinforcement learning is a kind of machine learning and do not require big data but learns based on its own producing data. Thus, it is possible to search the optimal action strategy or build model which accord with the objective function through learning the real time data. In addition, it is possible to apply the model to various environments because the model does not require any exact model of environment and therefore it is possible to control the traffic signal appropriately in real-time even in an urban signalized intersection where various geometric design and traffic volume exists. In specific, the agent observes the state of the environment and searches for the optimal action strategy to achieve specific objective, by selecting various actions. In this research, the environment, the state of the environment (S_t), the agent, the action (A_t) is set each to signalized intersection, delay time and emission of each movement, traffic signal control system, and signal phase. The model gets current state of the network, changes the signal phase, observes and formulates changing in state of traffic flow for decision criteria of the optimal phase plan. The action (A_t) which accord with the signal phase consists of 8 phases. These phases are possible combinations of each movement. (combination of 4 bounds of East, South, West, North and two directions of Left Turn and Straight) In addition, the reward (R_{t+1}) which is a criterion for the agent's action decision is set to changing delay time and emission by movement after displaying the decided signal, and this is maximised by the model. The objective function for minimizing both delay time and emission formulated in this research are shown as Eq(2) and Eq(3).

$$\max R_{t+1} = S_{t+1} - S_t \tag{2}$$

$$\mathrm{S_t} = \begin{cases} \sum_i d_{ij,t}\,, & \max d_{ij,t} \geq 50 \\ \sum_i e_{ij,t}\,, & \max d_{ij,t} < 50 \end{cases} \tag{3}$$

i is each vehicle in movement *j*; *j* is each movement in network; *t* is time step pf simulation, $d_{i\,j;\,t}$ is the delay of each vehicle *i* in movement *j* of time step *t*

Eq(2) represents the reward function for deciding the action (signal phase) through comparing between before and after the decided signal phase expresses. Eq(3) represents the state of the environment, and the state is different by the delay time of vehicle in the network. It is for considering the trade-off relationship between the delay time and emission rate of vehicle, caused by traffic signal control. The decision condition, 50 s, is a value of LOS (Level of Service) C proposed in KHCM (2013, Korea Highway Capacity Manual). In the reinforcement model, the value of the state is evaluated by the state-action pair which selected before and its observed reward. The agent observes the state of the environment (a delay time and an emission of each vehicle) and selects the action (signal phase) from the experience (cumulated data set of the state-action pair and reward value). In this process, the correct evaluation function and selecting function (control function) are required. Thus, the agent's observing and selecting process is repeated until the evaluation function and the control function are optimised. The update rule of evaluation and control function is shown as Eq(4).

$$Q_1(S_t, A_t) \leftarrow Q_1(S_t, A_t) + \alpha \left[R_{t+1} + \gamma Q_2 \left(S_{t+1}, \underset{a}{\operatorname{argmax}}\, Q_1(S_{t+1}, a) \right) - Q_1(S_t, A_t) \right],\ a \in A_t \tag{4}$$

In Eq(4), α is the learning rate, γ, discount factor, a is an action and Q (S_t, A_t) is the Q-function for evaluation and control. Unlike traditional Q-function, double Q-function is adopted as optimization function. The double Q-function separates evaluating the value of possible actions and selecting the optimal action and implements them by each Q-function to decrease the error which could be arisen at calculating. In Eq(4), γ represents the discount factor for adjusting between pursuing short-term and long-term reward. In addition, for updating the Q-function, it requires a large computing resources and times. This problem makes it harder to optimizing the Q-function, as the state-action pair becomes more complex. Because microscopic value, the delay time and emission value of vehicles, is adopted as the state explanation, more efficient method for optimizing the Q-function is required. In this point, the Deep learning method can be adopted. The Q-function combined with Deep learning is called Q-Network, and loss function for approximating the Q-Network to the optimal are shown as Eq(5) and Eq(6).

$$\min L_i(\theta_i) = \mathbb{E}_{s,a\sim\rho(\cdot)} \left[\left(y_i - Q(s, a; \theta_i) \right)^2 \right] \tag{5}$$

$$y_i = \mathbb{E}_{s'\sim\varepsilon} \left[r + \gamma \max_{a'} Q(s', a'; \theta_{i-1}) | s, a \right] \tag{6}$$

y_i is the target function and $\rho(s, \mathrm{a})$ is the possibility distribution of state s and action a. The Deep Neural Network method is applied for approximating the Q-Network to the optimal by minimizing $\mathrm{L_i}(\theta_i)$ in Eq(5), with some methods for relaxing the over fitting problem and efficient learning such as back propagation, drop out, and Xavier initializer.

3. Performance and Evaluation

3.1 Simulation configuration

The proposed model was implemented in microscopic traffic simulator - Vissim. A single intersection where major and minor streets intersect is constructed. Description of network is presented in Figure 1.

To evaluate developed traffic signal control system, comparison with an existing system was implemented. Two different scenarios are set for comparison, scenario 1 is a pre-timed signal controller, which optimises delay and scenario 2 is deep reinforcement learning-based signal controller. Both controllers optimise the delay time of vehicle, and all other conditions for simulation are same with the proposed signal control system.

3.2 Training

In 2,000 and 500 simulation steps, each of both models, the delay-optimizing model and the emission-optimizing model was trained every 15 steps. At early steps of simulation, displayed signals are selected randomly, but as training of models goes on, the models progress to their objects and selecting the signals by their own objects. The training results of both models are presented in Figure 2. In training the both models, there were some fluctuation of reward caused by exploration of the agent for finding the optimal signal phase. However, both

models got stable rewards as training goes on. This is because the definition of reward is set to the difference between before and after the traffic signal displayed. The training results indicates that each model makes each network condition stable.

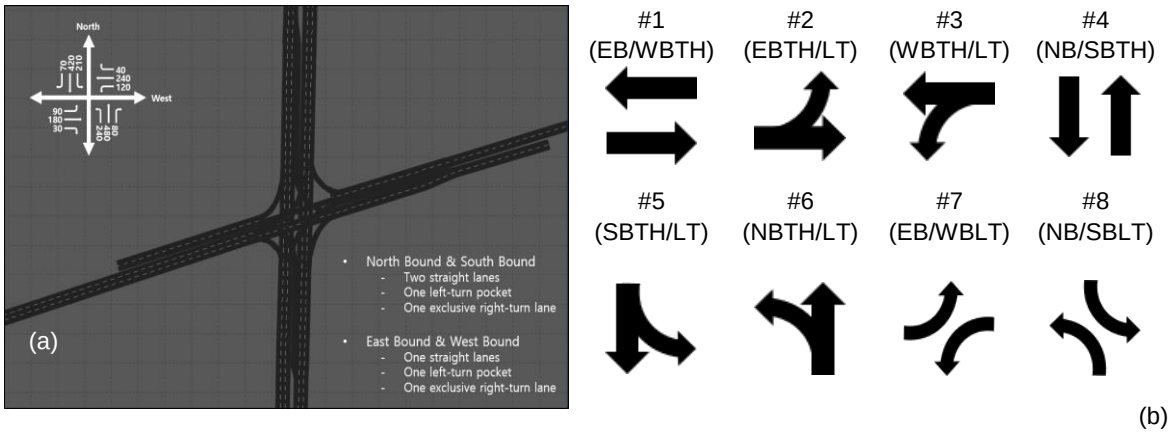

Figure 1: (a) Toy network for learning and evaluating (b) Available phases constructed by combining movement flow

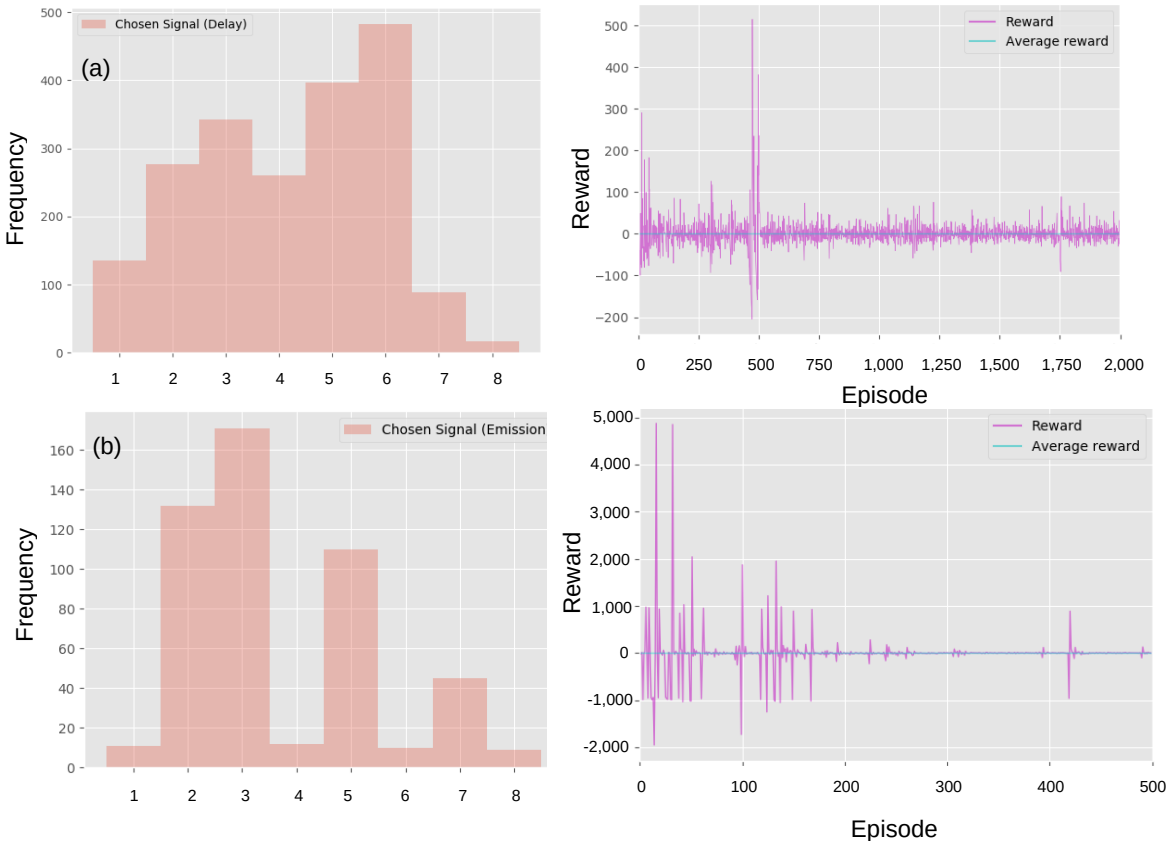

Figure 2: Training results (a) Delay-optimizing model (b) Emission-optimizing model

3.3 Evaluation

After training, both models were combined and evaluated. In specific, the model controls the traffic signal alternatively by whether the delay time of any vehicle in network exceeds 15 s. For evaluation, the model of scenario 1 controls the traffic signal by the optimised result through Passer V, the signal optimizing package. The model of scenario 2 is the single delay-optimising model, same as which is used in training. As a result, three models were evaluated for 200-time steps. Results of evaluation are present in Table 4. The first two

indices mean the obtained reward when the traffic signal controlled by each single model and 'Average obtained reward' is a real reward when the traffic signal is controlled by two models alternatively.

Table 4: Evaluation results of three models

	Proposed model	Scenario 1	Scenario 2
Average decreased delay (s/veh)	0.107453	0.021739	0.108585
Average decreased HC emission (mg/veh)	0.014195	-0.01743	-0.02611
Average obtained reward	20.59607	-	-

4. Conclusion

In this study, Double Deep Q Network is used in the V2X environment to perform reinforcement learning to enable agent's exclusive traffic signal operation. The objective of the study was to minimise the delay time and emission rate of vehicles by collecting and processing the information of each individual vehicles in real-time using COM interface feature of VISSIM. By combining pre-trained delay and emission optimizing model, it was possible to control the traffic signal accord with two different objectives, minimising vehicle's delay and emission rate, in the real-time condition of the network and surpassed traditional traffic signal control method. As stated in the evaluation section, our proposed model showed an average decreased delay about 0.107 seconds per vehicle, and this is about 4.9 times better than the traditional signal control method. Although average decreased delay of the proposed model was similar level to two delay-minimizing model, the proposed model showed decreased HC emission rate of about 0.014mg per vehicle, and other models showed increasing HC emission rate. This result shows that the multi-objective model optimizing different indices which affects traffic flow could be achieved through proposed method, such as combining different models presented in this study. By this, more factors which impacts signalized flow could be investigated with appropriate learning method and objective. However, more studies are needed to analyse the effect of proposed system, in more various road environment such as different proportion of V2X vehicles. For preparing oncoming V2X technology and smart green city, the proposed method of traffic signal control is expected to contribute on minimizing the emission occurred on road environment.

References

Ahn K., Rakha H., Trani A., Aerde M. V., 2002, Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels, Journal of Transportation Engineering, 128(2) 182 – 190.

Chitradevi S., Gayathri K., Anbarasi A., Karthika C., 2013, Model based traffic light control for reducing CO_2 emissions in vehicles, International Journal of Engineering Research & Techonology, 2(4) 154 – 159.

Dippold M., Hausberger S., Furian N., Haberl M., Hauser J., Stutzle T., Niebel W., 2015, Guideline for emission optimised traffic light control, Cooperative Self-Organizing System for Low Carbon Mobility at Low Penetration Rates, European Commission, Brussels, Belgium.

Koonce P., Rodegerdts L., Lee K., Quayle S., Beaird S., Bonneson J., Tarnoff P., Urbanik T., 2008, Traffic Signal Timing Manual, Federal Highway Administration, United States Department of Transportation, Washington DC, United States.

Li C., Shimamoto S., 2011, A real time traffic light control scheme for reducing vehicles CO_2 emissions, Consumer Communications and Networking Conference (CCNC), IEEE, 855 – 859.

LV J., Zhang Y., Zietsman J., 2013, Investigating emission reduction benefit from intersection signal optimization, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 17(3), 200 – 209.

Ministry of Land, Transport and Maritime Affairs, 2013, Korea Highway Capacity Manual, Ministry of Land, Transport and Maritime Affairs, Seoul, Republic of Korea.

Mnih V., Kavukcuoglu K., Silver D., Rusu A. A., Veness J., Bellamare M. G., Graves A., Riedmiller M., Fidjeland A. K., Ostrovski G., Petersen S., Beattie C., Sadik A., Antonoglou L., King H., Kumaran D., Wierstra D., Legg S., Hassabis D., 2015, Human-level control through deep reinforcement learning, Nature, 518(7540), 529 – 533.

Park B. B., Yun I., Ahn K., 2009, Stochastic optimization for sustainable traffic signal control, International Journal of Sustainable Transportation, 3(4), 263 – 284.

Pol E. V. D., 2016 Deep reinforcement learning for coordination in traffic light control, MS. Thesis, University of Amsterdam, Netherlands.

Roess R., Prassas E., McShane W., 2011, Traffic Engineering 4th ed., Pearson Higher Education, Inc., Upper Saddle River, New Jersey, United States.

Sutton R. S., Barto A. G., 1998, Reinforcement learning: An introduction, Cambridge, The MIT Press, Cambridge, United Kingdom.