

Ramp Metering Control Based on the Q-Learning Algorithm

*Edouard Ivanjko¹, Daniela Koltovska Nečoska², Martin Gregurić¹,
Miroslav Vujić¹, Goran Jurković¹, Sadko Mandžuka¹*

¹Faculty of Transport and Traffic Sciences, University of Zagreb, Vukelićeva 4, HR-10000 Zagreb, Republic of Croatia

²Faculty of Technical Sciences, Bitola, St. Kliment Ohridski University, Makedonska falanga 33, 7000 Bitola, Republic of Macedonia

Emails: edouard.ivanjko@fpz.hr, daniela.koltovska@tfb.uklo.edu.mk, martin.gregurich@gmail.com, miroslav.vujic@fpz.hr, goran.jurkovic@fpz.hr, sadko.mandzuka@fpz.hr

Abstract: Modern urban highways are under the influence of increased traffic demand and cannot fulfill the desired level of service anymore. In most of the cases there is no space available for any infrastructure building. Solutions from the domain of intelligent transport systems are used, such as ramp metering. To cope with the significant daily changes of the traffic demand, various approaches with autonomic properties like self-learning are applied for ramp metering. One of these approaches is reinforced learning. In this paper the Q-Learning algorithm is applied to learn the local ramp metering control law in a simulation environment, implemented in a VISSIM microscopic simulator. The approach proposed is tested in simulations with emphasis on the mainstream speed and travel time, using a typical on-ramp configuration.

Keywords: Ramp metering, Q-Learning, autonomic systems, artificial intelligence, microscopic simulation.

1. Introduction

In recent decades a significant increase in the traffic demand has occurred. This trend is especially present in densely populated areas where daily traffic congestions occur regularly during rush hours. Congestion is an important issue in road traffic because it can significantly reduce the traffic users' Level of Service (LoS). Most significant negative impacts of the reduced LoS are delays in goods delivery, in public, freight and private transport, etc. The delays induced by traffic

congestion cause time losses for the drivers and passengers, as well as increased fuel consumption.

Considering the above mentioned facts, it is important to develop efficient highway management control methods in order to mitigate the congestion and restore the originally planned LoS. Such traffic management control methods are considered under the scope of Intelligent Transport Systems (ITS). Systems from the ITS domain are functionally built as a superstructure of several transportation control systems based on advanced optimization of the transport processes with the use of information-communication infrastructure and devices. Today autonomic properties, such as self-learning, self-adaptation, self-configuration, etc. are also added.

The area where the on-ramp and mainstream flows are actually coming in interaction is known as a downstream bottleneck. An intense downstream bottleneck occurs as a consequence of vehicles platoon entry from on-ramps and leads to traffic overspill on the highway or to peaks in the traffic demand [3]. The ramp metering control method can significantly improve the total travel time of such an urban highway [5].

In this paper the technique of Reinforcement Learning (RL), e.g., the Q-Learning algorithm was applied to design an adaptive control strategy based on learning for local ramp metering control. The control strategy developed has been tested under typical traffic conditions in a VISSIM road traffic micro-simulator.

2. Ramp metering

Ramp metering uses special traffic signals at on-ramps to control the rate or size of the vehicle platoons entering the mainstream traffic according to the current traffic conditions [5]. Awareness of the current traffic conditions for a particular highway segment (traffic flow, speed and occupancy, etc.) is achieved by analyzing the real time data collected from the road sensors (inductive loops), traffic cameras, etc. These sensors are usually placed on the on-ramps and on the main road as presented in Fig. 1.

It is possible to divide the ramp metering algorithms in two major categories or strategies: local (or isolated) and coordinated. Ramp metering algorithms categorized as local, take into account the traffic conditions on a particular on-ramp and the nearby segment of the urban highway where they are applied. A drawback of the local ramp metering algorithms is their unawareness of the overall traffic situation on the entire controlled highway segment. ALINEA is the most widely used local algorithm and its core concept is to keep the downstream occupancy of the on-ramp at a specified level by adjusting the metering rate [4]. Ramp metering algorithms are categorized as coordinated if they take into account the overall highway traffic [2]. According to the overall highway traffic state, the metering rate for every on-ramp is adjusted. It is possible to divide the coordinated algorithms into cooperative, competitive and integrated algorithms [3]. The usual procedure for cooperative algorithms is to detect a bottleneck and enrol several upstream on-ramps to create virtual on-ramp queues. The competitive algorithms obtain during

the execution local and global solutions appropriate for the current traffic situation. The smaller value of these two solutions is selected. The integrated algorithms are based on an optimization engine with defined boundaries and a goal that has to be achieved during the control period [10]. In this paper only local ramp metering is examined.

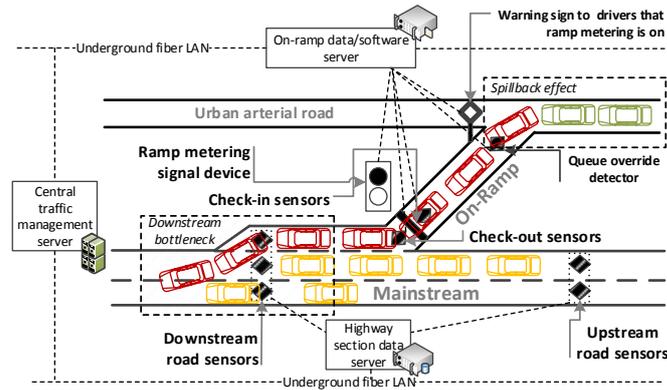


Fig. 1. Downstream bottleneck and spillback effect location with ramp metering infrastructure

3. Application of Q-Learning in ramp metering control

Standard ramp metering algorithms like ALINEA cannot cope with a wide range of traffic demand changes. Therefore, the researchers started to develop and apply various intelligent control approaches based on learning. Learning to act in ways that are rewarded is a sign of intelligence [7]. RL is one of the basic techniques of the Intelligent Agent (IA) technology. The learner or decision maker is named agent and everything it interacts with is the environment. The agent has a set of sensors to observe the state of the environment, and to perform a set of actions in order to change the state of the environment. The most important characteristics of the agent are the trial and error search, and delayed reward. The learner or an autonomous agent that senses its environment or acts in it can so learn through trials to select the optimal action or actions which lead to the largest reward.



Fig. 2. Agent-environment Interaction

For a more accurate presentation of the interaction we here assume that the agent and the environment communicate in all sequences of discrete time steps: $t = 0, 1, 2, \dots$. At each time step t the agent receives some representation of the state of the environment $s_t \in S$, where S is the set of possible states. In accordance with that, action $a_t \in A(s_t)$ is chosen, where $A(s_t)$ is a set of actions which are available in the state s_t . One step later, as a consequence of its action, the agent gets a numerical reward, $r_{t+1} \in R$ and finds itself in a new state, s_{t+1} . The agent obtains a reward or a

penalty in order to induce the desirability of the final state [11]. Fig. 2 shows the agent-environment interaction. The transition from one to another state is shown as

$$(1) \quad S_0 \xrightarrow[r_0]{a_0} S_1 \xrightarrow[r_1]{a_1} S_2 \xrightarrow[r_2]{a_2} \dots S_i \xrightarrow[r_i]{a_i} \dots,$$

where S_i is the state at the time step i , a_i is the possible action available in each state at the time step i , r_i is the reward which the agent receives at the time step i for taking action a_i . One of the most significant achievements in RL was the development of the Temporal Differences Off-Policy Algorithm known as Q-Learning. The ramp metering control strategy developed in this research is performed by an agent.

3.1. Q-Learning algorithm

Q-Learning provides the agent with an opportunity to learn the control policy. For non-deterministic environments the Q-function has been redefined as an expected value $\hat{Q}_n(s, a)$ from a previously defined value for a deterministic case. By applying the learning rule:

$$(2) \quad \hat{Q}_n(s, a) \leftarrow (1 - \alpha_n) \hat{Q}_{n-1}(s, a) + \alpha_n [r + \gamma \max_{a'} \hat{Q}_{n-1}(s', a')],$$

where the learning rate is

$$(3) \quad \alpha_n = \frac{1}{1 + \text{visits}_n(s, a)},$$

\hat{Q}_n still converges to Q^* [9] where Q^* is the optimal action value function, $\hat{Q}_n(s, a)$ is the expected value of the previously defined value for a deterministic function case for action a and state s and $\hat{Q}_{n-1}(s', a')$ is the expected value of the previously defined value for the new action a' at the next state s' . The parameter γ is the discount rate in the range of $0 \leq \gamma \leq 1$, α_n is the learning rate, (s, a) is the updated state and action during n iterations, and $\text{visits}_n(s, a)$ is the total number of visits for this pair of state-action until the n -th iteration.

3.2. Adaptation for local ramp metering

The proposed process of adaptive control strategy development for local ramp metering control is composed of three steps which are described in the continuation with more details. During the first step, the development of a model is conducted. The second step includes the design and development of the Intelligent Agent (IA). The third step is related to the learned strategy testing and evaluation.

Step 1. Development of a model

Firstly it is necessary to determine the *set of states*. The selection of appropriate variables to describe the traffic process is crucial. Three sets of states S

are defined: (1) a current phase, (2) a category of average mainstream speed and (3) a category of on-ramp queue. Description of the states S can be found in Table 1.

Table 1. Description of all states S

States S	Values	Description
Phases	1	Represents “all green” phase with fixed duration of 3 s (one vehicle per green strategy)
	2	Represents “all red” phase calculated by ramp metering algorithm (extension of current phase duration)
AverageSpeedClass	0	Downstream speed is between 0 and 70 km/h
	1	Downstream speed is between 71 and 99 km/h
	2	Downstream speed is larger than 100 km/h
AverageQueueClass	0	On-ramp queue length is between 0 and 4 vehicles
	1	On-ramp queue length is between 5 and 7 vehicles
	2	On-ramp queue length is larger than 8 vehicles

Based on the information related to the detected state, the control agent takes an appropriate action. The first possible action is denoted by the value 1. The mentioned action suggests that it is necessary to stay at the current traffic light phase. The second possible action indicates the necessity to change the current traffic light phase and it is denoted by the value 2. It has to be noticed here that for ramp metering only the green and red traffic light phases are used.

The rewarding function is the second key element for the agent. The reward is a function that depends on the system’s state and the action taken. Each action, derived by the agent, influences the environment. So upon completion of an action, the environment is in a new state. The agent is rewarded if the action is good. A reward for a particular traffic solution is added in case the on-ramp queue category has a value 0 or/and 1. Additional reward is added in case the speed reaches category 0.

The Q -matrix (size 35×4) represents the knowledge base of our agent and it stores the memory of what the agent has learned via numerous trials [11]. This approach of describing the Q -function in a look-up table is simple to use. Definitions and values of the parameters α and γ are displayed in Table 2.

Table 2. Parameters α and γ values applied in the research

Parameter	Parameter definition and value
α	The point of changing α is changing the balance of the exploration/exploitation relationship that the algorithm is to be determined, whereby at the beginning the algorithm is set to do more search for the solution and later it is set to optimize the solution it has found (the value of α is 0.9)
γ	The parameter is responsible for the reward transfer. It determines the influence of the future rewards over the agent’s behaviour (the value of γ is 80)

Step 2. Design and development of the IA

In the design of the intelligent agent, tools following software has been used: microscopic traffic simulation VISSIM5.4-0.3 (Germ. **V**erkehr **I**n **S**tädten **S**IMulationsmodell), specialized library dedicated for communication between VISSIM and Visual Studio-VISSIM **COM** (Engl. **C**OMPONENT **O**BJECT **M**ODEL) and database management tool Microsoft SQL [11].

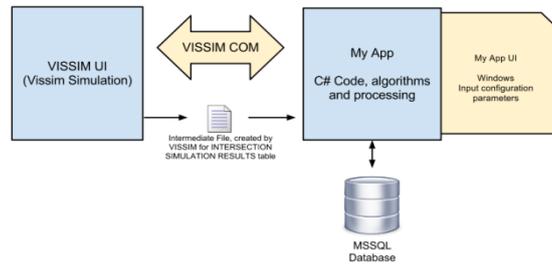


Fig. 3. The process of communicating and interaction among the main elements

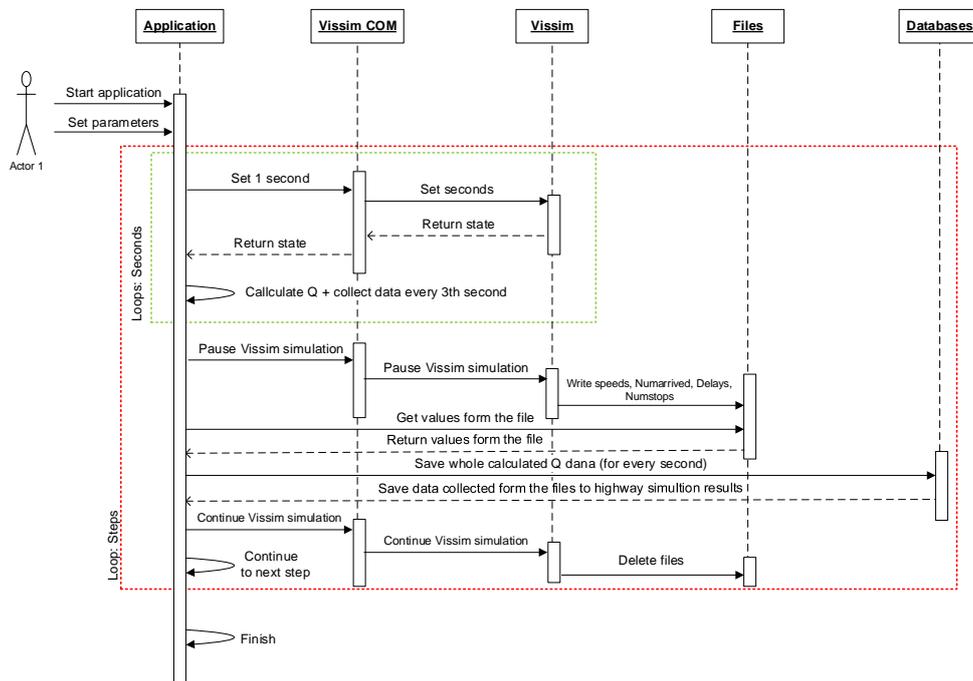


Fig. 4. Sequence diagram

The traffic demand has been created via VISSIM simulator's graphical interface. The number of vehicles is entered for every link within intervals of 5 minutes per peak hour (known/unknown demand). The vehicle arrivals are described by the Poisson distribution. To express the stochastic variations of the traffic flows as realistically as possible, a parameter used to initialize a random number generator is applied (*Random Seed*). In Fig. 3 the communication process and the interaction among the main elements is presented and in Fig. 4 the sequence diagram of the simulation setup is given.

4. Simulation results and evaluation

A highway model with one on-ramp is created in order to test the proposed local ramp metering approach based on RL. The created model contains additional detectors so that all relevant traffic parameters could be monitored. The simulation

results with and without the use of ramp metering are compared. In this paper a Q-Learning algorithm with its constant parameters set is applied during only one simulation step. Q-Learning algorithm is also applied with constant parameters in four simulation steps whereby the algorithm parameters are dynamically changed at every new simulation step.

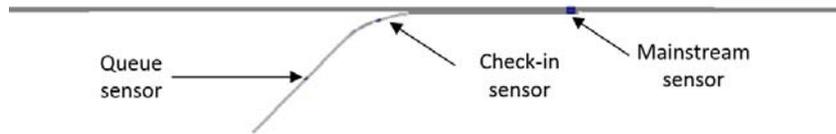


Fig. 5. Graphical representation of the highway model

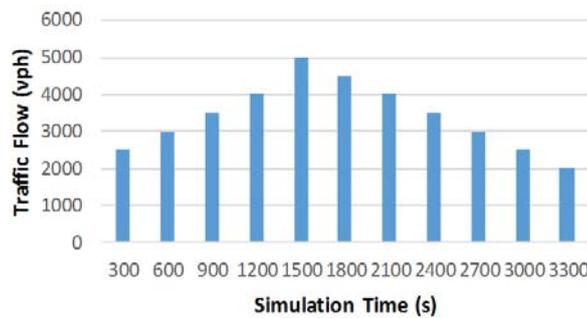


Fig. 6. Traffic flow values for every 5 minutes interval

4.1. Urban highway model

The model of the highway used in order to test the application of the Q-Learning based ramp metering is 1458 m long. It contains one on-ramp which is located 900 m from the starting point of the model. Furthermore, the model contains one signal head and three groups of traffic sensors.

Table 3. Comparative analysis between Q-Learning based ramp metering algorithm and no ramp metering applied

Parameter	No ramp metering (1 step)	Q-Learning algorithm (1 step)	Q-Learning algorithm (4 steps)
Mainstream travel time (h)	80.24	72.1	69.8
On-ramp travel time (h)	60	100.5	112.2
Average downstream speed (km/h)	63.4	70.5	72.1
Average on-ramp speed (km/h)	48.7	29.1	28.5
Average Speed of whole highway (km/h)	64.31	64.83	74.13
Total delay (h)	31.84	29.55	27.28
Total travel time (h)	80.94	78.05	75.82

First a traffic (check-in) sensor is placed at the entrance of the on-ramp. The second traffic (queue) sensor detects the queue length and it is located 150 m before the previously mentioned sensor. A third traffic (mainstream) sensor has the task to obtain the mainstream traffic data and it is placed 188 m downstream of the on-ramp. Graphical representation of the highway model which is used for simulation testing is presented in Fig. 5. The duration of the simulation is set to 3300 s. The

on-ramp flow was set to 450 vech/h being constant during the whole simulation. The mainstream in-flow values of the simulation model are constantly changed during every 5 minutes interval. The traffic flow values for every 5 minutes interval in the simulation run are presented in Fig. 6.

4.2. Obtained results

The use of ramp metering based on the Q-Learning algorithm (the one step and four steps version) is compared to the no ramp metering scenario. Results of the mentioned comparative analysis are presented in Table 3. In case when the Q-Learning algorithm is used only during one step, parameter α has a constant value of 0.9 while parameter γ has the value of 90. During the 4 steps, the Q-Learning algorithm updates the Q-Learning matrix with values from previous steps. But at every next step parameter α decreases its values by 0.2 and parameter γ by 15 with respect to the previous step. The influence of the future rewards over the agent's behaviour is decreased at every following step. It can be concluded that for every following step the results of the Q-Learning main equation have smaller effects on finding the optimal solution.

From Table 3 it can be concluded that Q-Learning algorithm can produce better values of the total travel time in comparison with the situation with no ramp metering. In case when Q-Learning algorithm is conducted during 4 steps, longer on-ramp queues are created. This implies a shorter mainstream and total travel time. The negative impact of this solution is a longer on-ramp travel time.

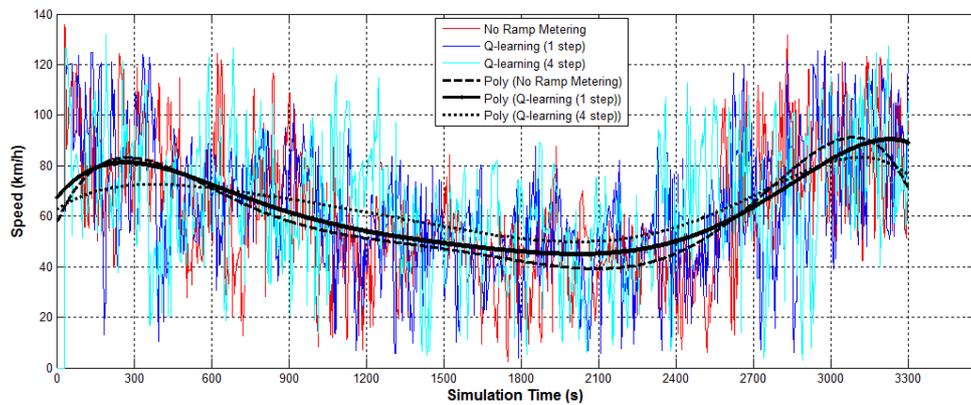


Fig. 7. Mainstream speed data analysis

In Figs 7 and 8 the speed and density are presented in 3 seconds intervals. Polynomial fit of 6th order has been done and presented in Figs 7 and 8 for the sake of a clear graphical representation of the monitored traffic parameters change trends. In the graphs mentioned, a typical behaviour of the simulated highway segment with respect to the average speed and density is given.

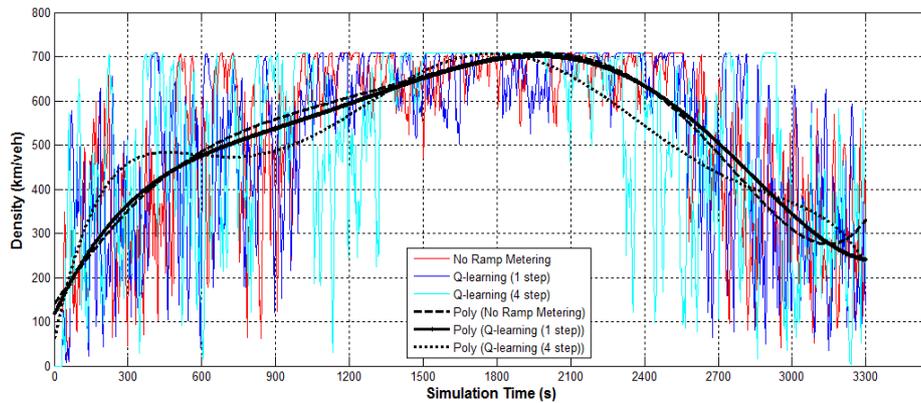


Fig. 8. Mainstream density data analysis

4.3. Discussion of the results

The first simulation results obtained show improvement of the mainstream travel time compared to the traffic situation without ramp metering applied. Learning IA based on Q-Learning is especially efficient during periods of a high traffic demand. Additionally, it is possible to conclude that the on-ramp travel time is larger and the on-ramp average speed is lower with the use of the learning IA compared to the situation without ramp metering. These negative results can be explained by the longer queues at the on-ramp induced by the IA in order to increase the mainstream speed. The longer on-ramp travel time is justified only in case when the value of the total travel time is not lower compared to the situation when the ramp metering is not applied.

Running the Q-Learning algorithm during four steps under different parameters has produced improvements compared to the case with only one step. Application of the described learning IA is not advisable in traffic scenarios characterized by low traffic load on the mainstream. In such cases it is better to turn off the ramp metering since no traffic demand related restrictions for vehicles entering the mainstream exist.

5. Conclusion

In this paper the Q-Learning algorithm was applied for local ramp metering. The Q-Learning algorithm was used to train an IA. Motivation for this research was to implement autonomic learning properties into local ramp metering by adjusting the parameters of the Q-Learning algorithm using an IA. The approach proposed was tested in simulations using a VISSIM based micro-simulation setup. An application that connects with a VISSIM micro-simulation model and saves all relevant traffic data in a database has been applied. Several types of the travel time (total travel time, on-ramp travel time and mainstream travel time) are analyzed as strategy efficiency measures. Additionally, the overall average speeds and mainstream density are also included in the analysis. The simulation results given in this paper are promising. A learning based IA for local ramp metering can be successfully

applied. The advantages, compared to the situation without ramp metering, are most noticeable in the mainstream travel time and average mainstream speed values improvement. According to the obtained results, the learning IA is not adequate in traffic scenarios with low mainstream speeds where ramp metering is usually switched off.

Future work will include tuning of the applied reward system. Additionally, cooperation between several on-ramps and additional highway control systems like variable speed limit, based on a learning IA will be also examined.

Acknowledgements: The research reported in this paper is partially funded by the European Union on FP7 Collaborative Project “Intelligent Cooperative Sensing for Improved Traffic Efficiency – ICSI” (FP7-317671) and by the EU COST action TU1102 “Towards autonomic road transport support systems”. The publishing of this paper is supported by FP7 project 316087 ACOMIN “Advance Computing and Innovation”. The authors wish to thank Nikola Bakarić for his valuable comments during writing of this paper.

References

1. Ghods, A. H., A. R. Kian, M. Tabibi. A Genetic-Fuzzy Control Application to Ramp Metering and Variable Speed Limit Control. – IEEE International Conference on Systems, Man and Cybernetics, ISIC’2007, 10 October 2007, Montreal, Quebec, Canada, pp. 1723-1728.
2. Hegyi, A., B. De Schutter, H. Hellendoorn. Model Predictive Control for Optimal Coordination of Ramp Metering and Variable Speed Limits. – Transportation Research, Part C, Vol. **13**, 2005, No 3, pp. 185-209.
3. Papageorgiou, M., C. Diakaki, V. Dinopoulou, A. Kotsialos, Y. Wang. Review of Road Traffic Control Strategies. – Proceedings of IEEE, Vol. **91**, December 2003, No 12, pp. 2043-2067.
4. Papageorgiou, M., H. Hadj-Salem, J. M. Blosseville. ALINEA: A Local Feedback Control Law for Onramp Metering. – Transportation Research Record, Vol. **1320**, 1991, pp. 58-64.
5. Papageorgiou, M., A. Kotsialos. Freeway Ramp Metering: An Overview. – In: Proc. of IEEE 5th Intelligent Transportation Systems Conference, Oakland, USA, 2002, pp. 271-281.
6. Papamichail, I., A. Kotsialos, I. Margonis, M. Papageorgiou. Coordinated Ramp Metering for Freeway Networks – A Model-Predictive Hierarchical Control Approach. – Transportation Research, Part C, Vol. **18**, 2010, No 3, pp. 311-331.
7. Sutton, R. S., A. G. Barto. Reinforcement Learning – An Introduction. Cambridge, Massachusetts, MIT Press, 1998.
8. Treiber, M. A. Kesting. Traffic Flow Dynamics. Berlin, Springer-Verlag, 2013.
9. Watkins, H. Learning from Delayed Rewards. PhD Thesis, University of Cambridge, 1989.
10. Gregurić, M., E. Ivanjko, S. Mandžuka. New Concepts for Urban Highways Control. – In: REAL CORP’2014. – Clever Plans for Smart Cities, Vienna, Austria, 19-20 May 2014, pp. 423-431.
11. Koltovska, D., K. Bombol. Intelligent Agent Based Traffic Signal Control on Isolated Intersections. – TEM Journal, Vol. **3**, 2014, No 3, pp. 216-222.