Application of Reinforcement Learning as a Tool of Adaptive Traffic Signal Control on Isolated Intersections

Kristi Bombol, Daniela Koltovska
UKLO: Department of Traffic and Transport
Faculty of Technical Sciences
Bitola, Republic of Macedonia
e-mail: kristi.bombol@uklo.edu.mk
e-mail:daniela.koltovska@uklo.edu.mk

Kostandina Veljanovska
UKLO: Information Systems Management
Faculty of Administration and Information Systems Management
Bitola, Republic of Macedonia
e-mail: kostandina@rocketmail.com

Abstract-**A large number of intelligent systems have been improved and adapted to the flow changes as well as for effective management of the signal time changes.**

For a long time it was believed that the systems responding to real time traffic would enable significant benefits. However, numerous limitations have appeared such as the existence of the models with high level of detail precision, the uncertainty in predicting future traffic flows, the difficulty in arrival time estimation, the lack of self-adjusting mechanism.

The difficulties in optimising the signal control strategy and the importance of finding a solution to this problem have initiated a great number of new researches. The results highlight the artificial intelligence methods as a possible solution. It has been determined that the artificial intelligent systems achieve the same results that man does when performing cognitive tasks. These systems are characterized with the ability to accumulate and use knowledge, set a problem, learn, process, conclude, solve the problem and exchange knowledge.

The research presented in this paper proposes an adaptive signal control performed by a control agent able to adapt to an optimal policy by learning from the environment. The goal to be achieved is minimization of the delays in the system.

First, the problem of reinforcement learning has been set. The first computation results of the Q-learning application for adaptive traffic signal control are presented.

It is concluded that the results obtained are in favor of the adaptive signal control strategy compared to the fixed and actuated signal control.

Keywords-isolated intersection, fixed control, fully – actuated control, artificial intelligence, reinforcement learningng, Q learning, simulations, delay

I. INTRODUCTION

The Intelligent Transport Systems (ITS) used by the advanced technologies in the transport system are an approach widely dispersed as a solution to the traffic problems that the society faces.

The aim of the use of ITS is improvement of the traffic quality by avoiding traffic congestions, saving travelling time, improving the safety and the comfort of the drivers and the passengers, improving the transport and the exchange of goods and services, and improvement of the entire informative transparency. Once achieved, all of these will lead to a higher level of satisfaction of the user's needs and prosperity of the surrounding area. Yet, there are more areas in which ITS can be applied and the traffic management and control in cities make only one of them.

Traffic signal control strategies are the most commonly used for managing the traffic flow at different types of intersections in the cities. At a signalized intersection they operate in one of the three different control modes: pre-timed control, semi actuated control, and fully-actuated control (Wilshire et al., 1985) [7].

Over the past two decades, significant efforts have been made to develop efficient and practical real time strategies to control the traffic at intersections. Although the concept of these efforts is promising, there are still some difficulties. The need to predict the origin and destination traffic demand in real time, the inherent limits in modeling the complex traffic flows, and the lack of confident sensors impose further investigations [8].

The researchers are expected to find simpler methods. Those have to include directly measured traffic parameters in the determination of the number of vehicles allowed to pass (through traffic signals), without the prediction in real time. The artificial intelligence can offer a very different approach to solve the above mentioned problems. The artificial intelligent systems reach the same results that man does when performing cognitive tasks. These systems are characterized with the ability to accumulate and use knowledge, set a problem, learn, process, conclude, solve the problem, and exchange knowledge.

The machine learning is a field of artificial intelligence. The reinforcement learning is a technique within the machine learning. It is successfully applied in solving problems, e.g. operating the elevators and robot soccer games. It is also applied in modeling the supply chain, dynamic allocation of resources, predicting time series [3].

In this research the reinforcement learning was applied in development of adaptive signal control strategy at an isolated intersection.

The paper is divided into four parts. The theory of the reinforcement learning is presented in the first part. The system model, including definitions pertaining the state, action, and reward function are in the second part. Part 3 presents and discusses the simulation results. The conclusions follow in part 4.

The analysis of the results of the simulation showed significant improvements in comparison with those of the fixed and fully-actuated signal control.

II. ARTIFICIAL INTELLIGENCE

A. *Reinforcement learning (RL) theory*

The ability to learn is one of the characteristics that define intelligence. That is why machine learning is center point of the artificial intelligence.

The reinforcement learning is a subfield of the machine learning, learning what to do i.e. how to map the state into actions, how to maximize the numerical reward, and in which way [1]. It is a type of learning driven by interaction with the environment and directed to a goal. It is a technique that does not need monitoring and it is different from the methods of the supervised learning (e.g. neural networks) [1] [2].

The learner or decision maker is named *agent*, and everything it interacts with is named *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: trial and error search and delayed reward.

The learner or an autonomous agent that senses its environment or acts in it can learn through trials to select the optimal action or actions which lead to the highest reward.

For a more accurate presentation of the interaction we here assume that the agent and the environment communicate in each sequence of discrete time steps: t=0,1,2,… In 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 teacher provides a reward or a penalty in order to induce the desirability of the final state. Figure 1 shows the agent-environment interaction.

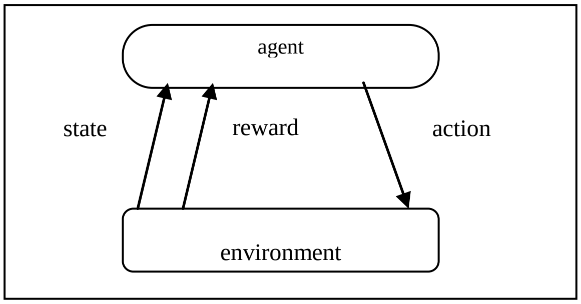

Figure. 1. AGENT-ENVIRONMENT INTERACTION

The transition from one to another state is shown as:

$$s_0 \xrightarrow[r_0]{a_0} s_1 \xrightarrow[r_1]{a_1} s_2 \xrightarrow[r_2]{a_2}$$

Where

S_i is the state in the time step i,

a_i is the possible action available in each state in the time step i,

r_i is the reward which the agent receives in the time step i for taking action a_i.

B. *Reinforcement learning elements*

The reinforcement learning has four basic elements, such as:

Policy – In each time step, the agent maps the presentation of the state with the probability of selection of each possible action. This mapping is called **agent policy** π_t, where $\pi_t(s,a)$ is the probability that $a_t=a$ if $s_t=s$. The reinforcement methods determine how the agent changes its policy as a result of its experience. The agent's goal, generally speaking, is to maximize the total reward it receives in long term.

Reward function – In the case of the reinforcement learning, the agent's goal is formalized in a sense of special signal, called **reward,** which is transmitted from the environment towards the agent. The reward is simply a number whose value differs from step to step. Informally, the agent's goal is to maximize the reward it receives. This means maximizing the reward that it receives in the moment as well as maximizing the cumulative reward that it receives in a long term.

Value function – Almost all reinforcement learning algorithms are based on the assessment of value functions – state functions (or state-action pairs) which assess how good a given state is for the agent to be in (or how good some action is to conduct in a given state). The term 'how good' is defined here in a sense of future rewards that can be expected, or more precisely said, in a sense of expected earning (compensation). Of course, the rewards that the agent expects to receive in the future depend on the action it will conduct. Accordingly, the value functions are defined in accordance with some policies.

As the agent learns directly from the interaction with the real environment, it does not need the **model of environment.** The agent can be taught in advance by means of a simulator. The use of a simulator helps the agent to avoid the unacceptable and degrading actions in the field. In a case of real time traffic control, this type of action can have a counter effect on the system in particular.

III. Adaptive Signal Control Strategy Design on Isolated Intersection

With the aim of collecting research data for this control strategy, simple network was created in a traffic micro simulator PTV Vision VISSIM COM (Fig. 2).

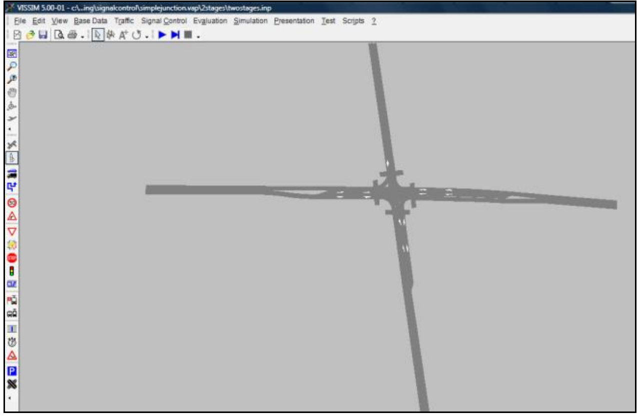
VISSIM 5.00-01 - c:\...ing\signalcontrol\simplejunction.vap\2stages\twostages.inp
File Edit View Base Data Traffic Signal Control Evaluation Simulation Presentation Test Scripts ?

Figure 2. INTERSECTION DESIGN

The input data consisted of:

- Data on the traffic network – created graphically
- Data on the traffic demand – entered through the graphical interface of the simulator as a profile of traffic demand in all zones – origins of traffic.

The data for measuring the efficiency of the control strategy were collected with the assistance of the functions inside the VISSIM and then, saved separately.

The functions were called in each time step. The time step in use was 1 second. The program for communication with the simulator, the collecting of the data, and the reinforcement learning algorithm were done in VBA.

The design process was done in the following steps:

- Creation of traffic network with graphical implementation of the intersection, creation of links, connectors, traffic signals, setting detectors
- Creation of traffic demand
- Gathering and analysis of the results

As it can be seen in Fig.2, there are four zones for generation of traffic demand, one for each approach. The name, the position and the dimensions of the detector were defined with the help of the graphical interface of the simulator. Each detector was linked to appropriate traffic signal.

Powerful tool from the Intelligent transport systems that uses the technique of the artificial intelligence, so-called **reinforcement learning** – the Q-learning approach was chosen to deal with the stochastic nature of the traffic and to develop adaptive signal control strategy at the intersection.

A. *Definiton of the RL elements for the problem researched*

The reinforcement learning elements were defined - states (*s*), actions (*a*), and rewards (*r*). The isolated intersection was two-phased and the number of states was eight with two possible actions each.

In the developing strategy for the intersection shown in Fig. 2 we set:

States:
[green][no gap][no occupancy]
[green][no gap][occupancy]
[green][gap][no occupancy]
[green][gap][occupancy]
[red][no gap][no occupancy]
[red][no gap][occupancy]
[red][gap][no occupancy]
[red][gap][occupancy]

Action set was defined as:
0 ... take action to change the phase
1 ... take action to continue the green phase

The choice of action for this research was made according to ε - greedy policy, where the best action is used with the possibility 1- ε, and the research action was chosen by random choice with probability ε.

Reward was defined as a total number of vehicles at the intersection.

The Q-learning approach was chosen to deal with the stochastic nature of the traffic and to develop adaptive signal control strategy at the intersection. The Q-learning algorithm [1] (control algorithm of the temporary difference (off-policy TD Control)) is shown in its procedural form in the following way:

```
Initialize $Q(s,a)$ arbitrarily
Repeat (for each episode):
    Initialize $s$
    Repeat (for each step of episode):
        Choose $a$ from $s$ using policy derived from $Q$ (e.g., - greedy)
        Take action $a$, observe $r$, $s'$
        $Q(s,a) \longleftarrow Q(s,a) + \alpha[r + \gamma \max_{\alpha'} Q(s',a') - Q(s,a)]$
    S←s';
until $s$ is terminal.
```

In the algorithm above *Q (s,a)* is the function of the action value, α is the learning percentage, *r* is the reward, γ is the discount rate, $Q(s',a')$ is the value function for the new action a' and the new state s'. S is the set of possible states.

IV. Simulation Results

In order to evaluate the proposed adaptive control strategy, the results obtained by the agent were compared with the results obtained by fixed and actuated signal control.

The testing was performed after sufficient number of iterations with different values of the conditions and after converging the Q-values. The number of iterations needed for convergence depends on the size of the state – action space. The simulation was run during the peak hour. The efficiency of the strategy was measured through delays.

In T.1 only a part of the results are shown. The comparison of average delays with the fixed signal control, actuated control, and Q-learning strategy is made.

TABLE I. Comparison of the average delays for fixed control, actuated control, and Q-learning strategy

Run	**Comparison of average delays for different types of signal control**	
	Type of Control	***Average Delay (sec)***
278	Fixed	0.8312433958
279	Fixed	0,6764412522
280	Fixed	0,9191551175
281	Fixed	0,5908772349
282	Fixed	0,6422748566
283	Fixed	1,1516582966
284	Fixed	0,6924495101
285	Fixed	0,6976267099
278	Actuated	0,232619598507881
279	Actuated	0,175925269722939
280	Actuated	0,192265138030052
281	Actuated	0,1535364985466
282	Actuated	0,159261509776115
283	Actuated	0,249249652028084
284	Actuated	0,15443129837513
285	Actuated	0,163399875164032
278	Q - learning	0,187686145305634
279	Q - learning	0,152055725455284
280	Q - learning	0,182069316506386
281	Q - learning	0,161372631788254
282	Q - learning	0,17327706515789
283	Q - learning	0,191062957048416
284	Q - learning	0,1720094711
285	Q - learning	0,1137311101

According to the modest simulation results obtained, one can expect some promising and encouraging conclusions. Namely, some reductions of average delays in favor of Q-learning strategy were perceived. However, this hypothesis cannot be scientifically proven at this time because of the small number of iterations and because of the simplicity of the isolated intersection design.

V. Conclusion

Although a great number of traffic control strategies has been developed and implemented so far, the field of adaptive signal control has still remained in the focus of many researchers over the past twenty years. The reason behind lies in their **numerous limitations such as the** uncertainty in predicting future traffic flows, the difficulty in arrival time estimation, the lack of self-adjusting mechanism.

This paper is an attempt to promote the application of adaptive signal control strategy at an isolated intersection by using the technique of artificial intelligence known as reinforcement learning. The research was conducted by using VISSIM micro simulator, and by direct programming of the functions in the simulator.

The adaptive traffic signal control strategy tested on isolated intersection provided encouraging results. The simulation results concerning the average delays indicated that the new approach was more efficient than the one with the fixed and actuated signal control.

Further evaluations of the strategy in different conditions are recommended and different settings of the elements of the Q learning agent have to be additionally developed.

References

[1]. Sutton, R.S., Barto, A.G., “Reinforcement Learning - An Introduction”. MIT Press, Cambridge, Massachusetts, 1998.

[2]. Russell, S., Norvi, P., “Artificial Intelligence: A Modern Approach”, Prentice Hall, 2009.

[3]. Artificial Intelligence in Transportation *Information for Application*, Transportation Research CIRCULAR, Number E – C 113, TRANSPORTATION ON RESEARCH BOARD *OF THE NATIONAL ACADEMIES*, January 2007

[4]. Abdullah B., Kitten L., “Reinforcement learning: Introduction to theory and potential for transport applications”, Can. J. Cave Eng. 30, pp. 981-991, 2003.

[5]. Abdullah, B., Pringle, R., Karakuls, G.J., “Reinforcement Learning for True Adaptive Traffic Signal Control”, ASCE Journal of Transportation Engineering, Volume 129, Number 3, pp278-285, 2003.

[6]. Yu, X.- H., Rocker, W.W., “Stochastic adaptive control model for traffic signal systems, Transportation Research Part C: Emerging Technologies, Volume 14, Issue 4, pp 263 – 282, Elsevier, August 2006

[7]. Shi, Z., “Principles of Machine Learning”, International Academic Publishers, Beijing, 1992.

[8]. Veljanovska, K., Bombol, K., “Choosing a new tool for adaptive control strategy design”, PhD Research day RESEARCH PROCEEDNIGS, May 9, 2007.

[9]. Spall, J. C., Chin, D., “Traffic -Responsive Signal Timing for System-Wide Traffic Control”, Transportation Research Part C, Vol. 5, No. 3/4, Elsevier Science Ltd, Great Britain, pp. 153-163,1997.

[10]. Wiering, M., Veenen,V, Jelle., Vreeken, J., Koopman, A., “Intelligent Traffic Light Control”, Institute of information and computing sciences, Utrecht University, Technical report UU-CS-2004-029.

[11]. Roozemond, D., Veer, P. V. D., “Usability of Intelligent Agent Systems in Urban Traffic Control”, Delft University of Technology, Delft, 1999.

[12]. Meystel, A & Messina E., “The Challenge Of Intelligent Systems”, Proceedings of the 15th, IEEE International Symposium on Intelligent Control (ISIC 2000), Rio, Patras, GREECE 17-19 July, 2000.

[13]. Shoufeng, L., Ximin, L., Shiqiang., “Q – Learning for Adaptive Traffic Signal Control Based on Delay Minimizations Strategy”, IEEE International Conference on Networking Sensing and Control, pp. 687 – 691, 2008.

[14]. Koltovska, D., Bombol, K., “Analysis of the Potentials of the Artificial Intelligence Techniques for Adaptive Signal Control”, ISEP 2009, R 2, Ljubljana, 2009.

[15]. “VISSIM 5. 00” User manual, PTV Planung Transport Verkehr AG, Germany, 2007.

[16]. Roca, V., “VISSIM COM” – User Manual for the VISSIM COM Interface – VISSIM 5.0, PTV Planung Transport Verkehr AG, Karlsruhe, 2007