International Balkans Conference on Challenges of Civil Engineering, BCCCE,
19-21 May 2011, EPOKA University, Tirana, ALBANIA.

Reinforcement learning approach for optimising traffic signal timings at isolated intersections

Cenk Ozan, Özgür Başkan, Soner Haldenbilen, Halim Ceylan

Transportation Division, Department of Civil Engineering, Pamukkale University, Denizli, Turkey

Abstract

One of effective ways to prevent congestion and delay on urban areas is signal control at intersections. Signal systems are operated according to state of intersections either isolated or coordinated signal systems. Many researches have been investigated to improve traffic signal systems based on delay minimization or capacity maximization throughput. Due to complexity of the system, new methods are needed to improve efficiency of signalization in a road network.

Signal setting parameters are usually obtained by minimizing total delay on an intersection. The delay is the key parameter which determines the level of service of an intersection. Delay is defined with two parts as an uniform and non-uniform. The uniform part of the delay is determined basically using conventional delay formulas. But the non-uniform part is not easily determined and cannot be represent due to the nature of the problem and randomness in arrivals.

In this study, **R**einforcement **L**earning **S**ignal **O**ptimizer (RLSO) is used to optimize signal timings in isolated intersection because of reflecting the effect of non-uniform part of delay. Reinforcement Learning (RL) which is an approach to artificial intelligence that emphasizes learning by the individual from its interaction with its environment. This contrasts with classical approaches to artificial intelligence and machine learning, which have downplayed learning from interaction, focusing instead on learning from a knowledgeable teacher, or on reasoning from a complete model of the environment. RL is learning what to do-how to map situations to actions-so as to maximize a scalar reward signal. The learner is not told which action to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them.

The aim of this paper is to minimize delay on intersections controlled by isolated signal system and to obtain operational parameters such as cycle time, green split rate. For this purpose, the RLSO is applied to an example intersection which has four approaches and three stages. The results of RLSO were compared with field observations. The results showed that the RLSO is able to optimize traffic signal timings on an intersection. The proposed model also holds promise for successful application to optimize traffic signal timings at isolated intersections according to delay minimization.

Keywords: *Signal optimization, reinforcement learning, isolated intersection.*

Introduction

In urban networks, traffic signals are used to control vehicle movements so as to reduce congestion, improve safety, and enable specific strategies such as minimizing delays, improving environmental pollution, etc [1]. Due to the increasing in the number of cars and developing industry, finding optimal traffic signal parameters has been an important task in order to use the network capacity optimally. Through the last decade, developments in communications and information technologies have improved the classical methods for optimising the traffic signal timings toward the intelligent ones [2].

Signal systems that control intersections are operated according to state of intersections either isolated or coordinated signal systems. Through the years, procedures for determining optimum signal timings have been developed and continuously improved. Early methods such as that of Webster [3] only considered a single signalized junction in isolation. Later, fixed time strategies were developed that optimized a group of signalized junctions using historical flow data (e.g., TRANSYT [4]). In some cities, real time traffic flow data has also been used for optimization in methods commonly referred to as demand responsive strategies (e.g., SCOOT [5]).

Traffic congestion is one of the main problems in urban areas, especially in metropolises. Traffic engineers or researchers try to solve the problems and also decrease traffic congestions. The problems result from different sources, such as vehicle arrival type, traffic control, and related parameters. On urban networks most of the total travel time is spent at intersections (delays). Hence, effective optimization of signalized intersections can significantly develop transportation network performance. Vehicle delay is one of the key parameters that is used for signalized junction design and level-of-service determination. Vehicle delay involves two parts as uniform and non-uniform by researchers. Uniform delay is calculated based on signal timings and traffic volumes. On the other hand, nonuniform delay is ascertained by considering the vehicle queue and random arrivals. While uniform delay is handled easily, determination of non-uniform delay has been a problem for researchers, especially for oversaturation cases. Many researches have been investigated to improve traffic signal systems based on delay minimization or capacity maximization throughput. Due to complexity of the system, new methods are needed to improve efficiency of signalization in a road network.

In this study, **R**einforcement **L**earning **S**ignal **O**ptimizer (RLSO) is used to optimize signal timings in isolated intersection because of reflecting the effect of non-uniform part of delay. The paper is organized as follows. Delay formulas are defined in the next section. Section 3 is about the RL algorithm. Numerical application is given in Section 4. Last section is about the conclusions.

Delay

Signalized intersection delay may be analysed into uniform component that consists of signal timings, and random plus oversaturation component that includes vehicle queuing, random arrivals and over-saturation cases of traffic flows [6]. The vehicle delay at signalized intersections has been defined as the stopping, acceleration, and deceleration delays.

The stopping delay includes the time the vehicles spend stopped during a red or green signal period at a signalized intersection. The acceleration delay of the vehicles may be defined as the time required to accelerate after the traffic signals turn to green. The deceleration delay is the time during which the vehicles decelerate while approaching a signalized intersection. In the evaluation of traffic signal control systems or signal design, the overall delay is the most commonly used value.

The Webster [3], Highway Capacity Manual (HCM) [7] or the Akcelik's [8] delay calculation methods have been preferred by traffic engineers for many years. In 1965, the Transportation Research Board (TRB) published the HCM, and it has subsequently been updated several times. In the HCM method, the average delay of vehicles is calculated based on a lane of an approach.

The measurements of queue lengths and vehicle delays in testing the predictions of time-dependent queuing models were studied [9]. The 1985 HCM delay formula with that of 1994 was compared [10]. Powell [11] proposed some correction factors representing the deceleration and acceleration delays of vehicles based on queuing to improve the 1997 HCM delay formula. Qiao et al. [12] developed a fuzzy logic model to simulate HCM delay formula. Dion et al. [13] compared various analytic models with microscopic simulation models and addressed delays at signalized intersection controlled in fixed-time and operated in a range of conditions. Murat and Baskan [14] studied vehicle delays using artificial neural networks (ANN). Traffic queues and delays at road junctions are reported [15]. Delay components are solved using coordinate transformation method [16]. Approximate mathematical expressions for delay components at signalized intersections were developed [17]. The delay has been calculated using the following equation proposed by Webster [3] in this study:

$$d = \frac{C(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)} - 0.65\left(C/q^2\right)^{1/3} x^{2+5\lambda} \tag{1}$$

where, d = average delay for vehicles on arm of the intersection, sec/veh; C = cycle time, sec; q = traffic volume, veh/h; λ = ratio of effective green to the cycle time; s = saturation flow, veh/h; x = degree of saturation. This formula is not valid if the degree of saturation is more than 1.

While Webster's formulas treat the delay due to individual cycle failures in the random delay term, they do not address the more significant issue of delay when demand exceeds capacity for a significant period of time. If demand continuously exceeds capacity, the overflow delay component continues to grow with time. Thus in the case of oversaturation, the duration of the oversaturation affects the overflow delay. The uniform delay component in this situation is a special case of first term Eq. (1) is given Equation (2) [18].

$$UD_o = \frac{C[1-\lambda]}{2} \tag{2}$$

where UD_o is the uniform delay component. The average overflow delay is given by Eq. (3):

$$OD_o = \frac{T}{2}\left[(q/c)-1\right] \tag{3}$$

where OD_o is the average overflow delay per vehicle, c is capacity, veh/h and T may be in seconds, minutes or hours. The total average delay per vehicle is given by Eq.(4):

$$d = UD_o + OD_o = \frac{C[1-\lambda]}{2} + \frac{T}{2}\left[(q/c)-1\right] \tag{4}$$

Reinforcement learning

The Reinforcement Learning (RL) problem is meant to be a straightforward framing of the problem of learning from interaction to achieve a goal. In RL, the learner and decision-maker are called the agent that it interacts with its environment. This interaction takes the form of the agent sensing the environment, and based on this sensory input choosing an action to perform in the environment. The action changes the environment in some manner and this change is communicated to the agent through a scalar reinforcement signal. The environment also gives rise to rewards, special numerical values that the agent tries to maximize over time. The agent and environment interact at each of a sequence of discrete time steps, $t = (0,1,2,3,....)$. At each time step t, the agent receives some representation of the environment's *state*, $s_t \in S$, where S is the set of possible states, and on that basis selects an *action*, $a_t \in A(s_t)$, where $A(s_t)$ is the set of actions available in state s_t. One time step later, the agent receives a numerical *reward*, $r_{t+1} \in R$, and finds itself in a new state, s_{t+1}. Fig. 1 shows the agent-environment interaction [19].

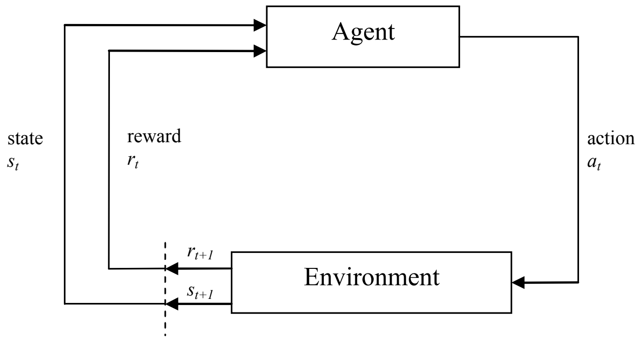

Figure 1 The agent-environment interaction in reinforcement learning

Q-learning is a model-free approach to RL that does not require the agent to have access to information about how the environment works. It works by estimating state-action values, the *Q*-values, which are numerical estimators of quality for a given pair of state and action [20]. The development of *Q*-learning is seen as one of the most important breakthroughs in RL. It uses the experience of each state transition to update one element of a table [19]. This table denoted *Q*, has an entry, $Q(s,a)$, for each pair of state, *s*, and action, *a*. Upon the transition s_t, s_{t+1}, having taken action a_t and received reward r_{t+1}. The *Q*-learning algorithm compromises of the *Q*-value, reflecting the value of an action *a* executed in a state *s* and selecting the best actions. The *Q*-values can be defined as follows:

$$Q(s,a) = r(s,a) + \gamma \times Q^*(s',a') \tag{5}$$

where $Q(s,a)$ is the *Q*-value of the state action pair (s,a) and $Q^*(s',a')$ is the best *Q*-value which can be obtained by selecting action a' in state s', which is the state resulting from executing action *a* in state *s*. $r(s,a)$ is the reward received when executing action *a* in state *s*. γ is the discounting factor, reflecting the weight assigned to future rewards [21]. The *Q*-table is populated in the course of the learning process in Fig. 2.

The learning process takes place in the course of a number of learning episodes. Each learning episode starts in a random state *s*, the agent selects and executes an action, receives the immediate reward and observes the next state. Based on this information, the agent updates the *Q*-value corresponding to this state-action couple according to the formula below:

$$Q_t(s,a) = (1-\alpha) \times Q_{t-1}(s,a) + \alpha \times \left[r(s,a) + \gamma \times \max_{a'} Q_{t-1}(s',a') \right] \tag{6}$$

where $Q_t(s,a)$ is the updated *Q*-value, $Q_{t-1}(s,a)$ is the *Q*-value previously stored in the *Q*-table and which needs to be updated and α is the step size parameter or learning rate of the algorithm and expresses the weight assigned to the "newly" calculated Q-value compared to the "old" saved estimate of the Q-value $Q_{t-1}(s,a)$ and γ is the discounting factor [22].

Initialize Q-values Repeat N times (N=number of the learning episodes) Select a random state s Repeat until the end of the learning episode Select an action a Receive an immediate reward $r(s,a)$ Observe the next state s' Update the Q-table for the state-action pair (s,a) according to update rule Set $s=s'$

Figure 2 Reinforcement learning process [22]

Numerical Application

The RLSO is applied to an example intersection which has four approaches and three stages. The basic flowchart of RLSO is given in Fig. 3. Isolated intersection and stage configuration can be seen in Figure 4. Corresponding traffic volumes for two cases are given in Table 1. Saturation flows are equally distributed to each lane are taken as 1800 veh/h. Initial set of signal timings for first iteration was random generated in RLSO according to given bounds of signal timings. In order to provide the constraint of cycle time for intersection, green timings can be distributed to the all signal stages as follows [23]:

$$\varphi_i = \varphi_{\min,i} + \frac{\varphi_i}{\sum_{k=1}^{m}\varphi_i}\left(C_i - \sum_{k=1}^{m} I_k - \sum_{k=1}^{m}\varphi_{\min,k}\right) \qquad i = 1,2,.....,m \tag{7}$$

where φ is green time (sec), φ_i is minimum green time (sec), m is the number of stages and I is intergreen time between signal stages.

In the RLSO process, the objective is to minimize delay, D, which has been calculated by Eq. (8).

$$\underset{\psi\in\Omega_0}{Min}\ D(\psi,q) = \sum_{k=1}^{m} d(\psi,q); \left\{ \begin{array}{ll} \text{if } x<1 & d = \dfrac{C(1-\lambda)^2}{2(1-\lambda x)} + \dfrac{x^2}{2q(1-x)} - 0.65\left(C/q^2\right)^{1/3} x^{2+5\lambda} \\ & \\ \text{if } x \geq 1 & d = \dfrac{C[1-\lambda]}{2} + \dfrac{T}{2}\left[(q/c)-1\right] \end{array} \right\} \tag{8}$$

$$\text{subject to} \quad \psi = (C,\varphi) \in \Omega; \left\{ \begin{array}{ll} C_{\min} \leq C \leq C_{\max} & \text{cycle time constraints for intersection} \\ & \\ \varphi_{\min} \leq \varphi \leq C & \text{green time constraints for each stage} \end{array} \right\}$$

where, D is total average delay for intersection (sec/veh); d is the average delay for each stage (sec/veh); ψ is signal setting parameters; C is cycle time (sec) for an intersection; φ is green time (sec) for each stage; q is traffic volume (veh/h) for each approach on an intersection and

c is capacity (veh/h) for each approach on an intersection. The signal timing constraints are given as follows:

$C_{\min}, C_{\max} = 40, 120\,\text{sec}.$ the bounds of cycle time for intersection

$\varphi_{\min} = 7\ \text{sec}.$ minimum green time for signal stages.

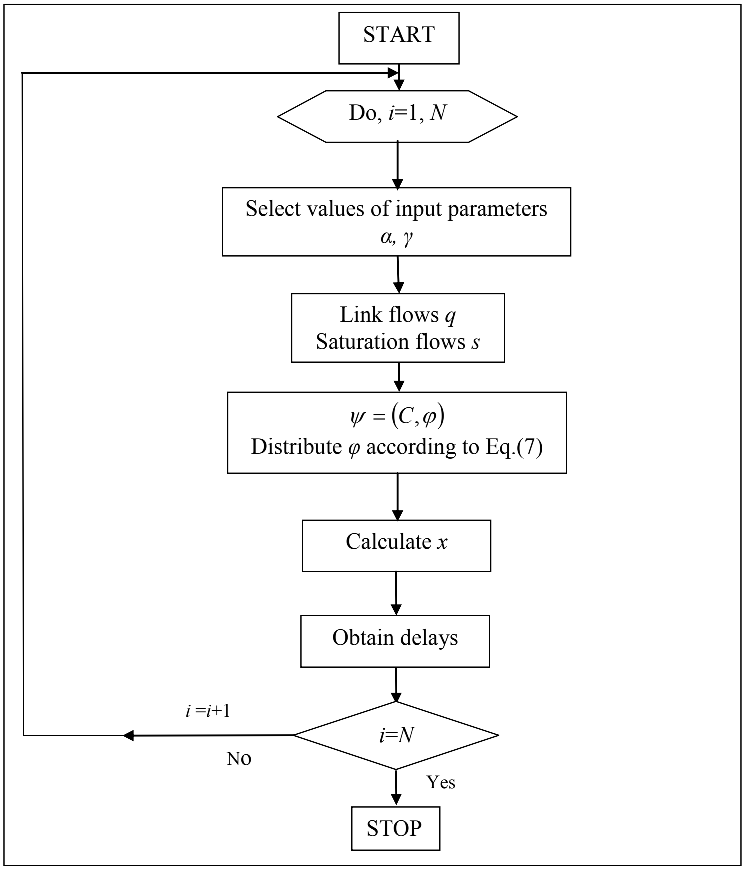

Figure 3 Flowchart of RLSO

Table 1 Traffic volumes for example intersection

Approach	Traffic volumes (veh/h)	
	Case 1	Case 2
A	200	90
B	400	750
C	250	100
D	350	550

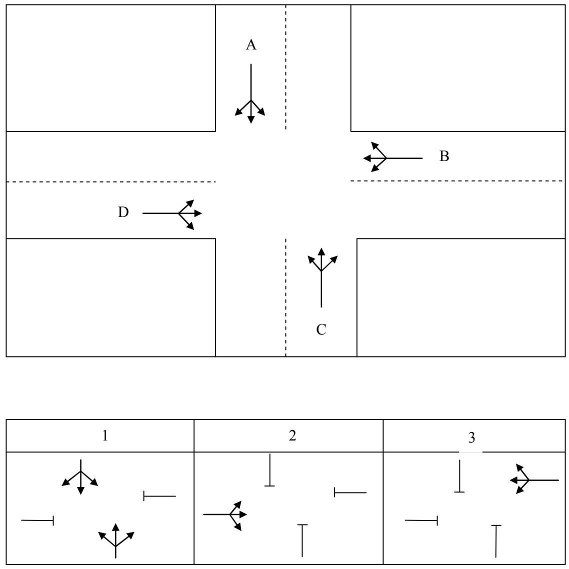

Figure 4 Example intersection

On the example intersection, delays were calculated with RLSO for each case. The obtained delays are given Table 2.

Table 2 Obtained delays on example intersection

	Stage	Optimum cycle time (sec)	Green time (sec)	Traffic volume (veh/h)	Degree of saturation (x)	Average delay sec/veh	
Case 1	1	44	10	250	0.61	19.27	84.98
	2		11	350	0.77	24.57	
	3		11	400	0.88	41.14	
Case 2	1	96	18	100	0.29	34.89	240.89
	2		29	550	1.01	41.63	
	3		37	750	1.08	164.38	

In case 1, delay was minimized according to given flows using RLSO. As shown in Table 2, the degree of saturation of each stage is less than 1 for this case. The delay of each stage was found 19.27, 24.57 and 41.14 sec/veh, respectively. At the end of process of RLSO, green timings are shared in stages in proportion to traffic volume as expected. Due to partially higher traffic volumes, the degree of saturation of stage 2 and 3 in case 2 is more than 1. In this case, intersection delay increased to 240.89 sec/veh. The green timings for this case were found 18, 29 and 37 sec, respectively.

Halley and İtfaiye intersections are located in Denizli, Turkey, which are controlled isolated signal system are taken into consideration to compare field studies and RLSO. They have four approaches and three stages. Traffic volumes, cycle times and green times of each stage of these intersections are given in Table 3 and 4 [14].

Table 3 Traffic volumes, cycle and green times for Halley intersection

Stage	Traffic volumes (veh/h)	Cycle time (sec)	Green time (sec)
1	385	92	30
2	222		16
3	252		34

Table 4 Traffic volumes, cycle and green times for İtfaiye intersection

Stage	Traffic volumes (veh/h)	Cycle time (sec)	Green time (sec)
1	595	112	35
2	170		35
3	90		30

On the Halley and İtfaiye intersections, delays were calculated using RLSO. The results were compared with the field studies as shown in Table 5. RLSO gives less cycle times and delay for these intersections when it is compared with field observations. Thus, RLSO may be used for optimising traffic signal timings on isolated intersections.

Table 5 Comparison of delays on Halley and İtfaiye intersections

Intersection	Stage	Cycle Time (sec)	Green times (sec)	Traffic volume (veh/h)	Degree of saturation (x)	Average delay sec/veh	
						RLSO	Observation
Halley	1	40	9	222	0.55	67.30	168.48
	2		10	385	0.86		
	3		9	252	0.62		
İtfaiye	1	50	11	90	0.22	115.72	129.87
	2		16	595	1.03		
	3		11	170	0.42		

Conclusions

In this study, RLSO which is based on RL is used to optimize signal timings on isolated intersection because of reflecting the effect of non-uniform part of delay. For this purpose, the RLSO was applied to an example intersection which has four approaches and three stages. Then, the RLSO was implemented to Halley and İtfaiye intersections located in Denizli, Turkey. The results of RLSO were also compared with field observations of these intersections. The results showed that the RLSO is able to optimize traffic signal timings on an intersection when it is compared with the field observations. The proposed approach also holds promise for successful application to optimize traffic signal timings on isolated intersections.

Acknowledgement

Scientific Research Foundation of the Pamukkale University with the project number 2010-FBE-063 is also acknowledged.

REFERENCES

[1] Teklu, F., Sumalee, A. and Watling, D. (2007) A genetic algorithm approach for optimizing traffic control signals considering routing. Computer-Aided Civil and Infrastructure Engineering, 22, 31-43.

[2] Baskan, O. and Haldenbilen, S. (2011) Ant Colony Optimization Approach for Optimizing Traffic Signal Timings. Ant Colony Optimization – Methods and Applications (Edited by Ostfeld, A.), InTech, Rijeka, Croatia, pp. 205-221.

[3] Webster, F. V. (1958) Traffic Signal Settings. Road Research Technical Paper No. 39, HMSO, London.

[4] Robertson, D. I. (1969) 'TRANSYT' method for area traffic control. Traffic Engineering and Control, 10, 276–81.

[5] Hunt, P. B., Robertson, D. I., Bretherton, R. D. and Winton, R. I. (1981) SCOOT—A traffic responsive method of coordinating signals. TRRL Laboratory Report 1014, TRRL, Berkshire, England.

[6] Ceylan, H., Haldenbilen, S., Ceylan, H. and Baskan, O. (2010) Development of delay models with quasi-Newton method resulting from TRANSYT traffic model. Journal of Scientific & Industrial Research Vol. 69, pp. 87-93.

[7] TRB, (1985) Special Report 209, Highway Capacity Manual, Transportation Research Board, National Research Council, Washington, DC, USA.

[8] Akçelik, R., (1981) Traffic Signals: Capacity and Timing Analysis Research Report 123. Australian Road Research Board, Melbourne, Australia.

[9] Kimber, R. M. and Daly, P. N., (1986) Time-dependent queuing at road junctions: observation and prediction. Transportation Research Part B 20B, 187–203.

[10] Prevedouros, P. D. and Koga, C. A., (1996) Comparison of 1985 and 1994 HCM: signalized intersection delay estimates, ITE J, 66, 26-30.

[11] Powell, J. L., (1998) Field measurement of signalized intersection delay for 1997 update of the highway capacity manual, Transportation Research Record, 1416, 79-86.

[12] Qiao, F., Yi P., Yang. H. and Devarakonda, S., (2002) Fuzzy logic based intersection delay estimation, Mathematical and Computer Modelling, 36, 1425-1434.

[13] Dion, F., Rakha, H. and Kang, Y. S., (2004) Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections, Transportation Research, 38B, 99-122.

[14] Murat, Y. S. and Baskan, O., (2006) Modeling vehicle delays at signalized intersections: Artificial neural networks approach, Journal of Scientific and Industrial Research, 65, 558-564.

[15] Kimber, R. M. and Hollis, E. M., (1979) Traffic queues and delays at road junctions, in TRRL Laboratory Report 909 (Transport and Road Research Laboratory, Crowthorne, Berkshire).

[16] Chiou, S.-W., (1998) Optimisation of area traffic control for equilibrium network flows, Ph D Thesis, University College London.

[17] Ceylan, H., Baskan, O., Ceylan, H. and Haldenbilen, S., (2007) Mathematical solutions of vehicular delay components at signalized intersections based on approximate calculation method, Journal of Engineering Science, 13, 279-288 (in Turkish).

[18] Roess, R. P., Prassas, E. S., and McShane, W. R. (1998) Traffic Engineering. Second Edition, Prentice Hall. Upper Saddle River, New Jersey.

[19] Sutton, R.S. and Barto, A.G. (1998) Reinforcement Learning – An Introduction, The MIT Press.

[20] Bazzan, Ana L.C., Oliveira, D. and Silva, B.C. (2010) Learning in groups of traffic signals. Engineering Applications of Artificial Intelligence, 23, 560-568.

[21] Vanhussel, M., Janssens, D., Wets, G. and Vanhoof, K. (2009) Simulation of sequential data: An enhanced reinforcement learning approach. Expert Systems with Applications, 36, 8032-8039.

[22] Kaelbling, L.P., Littman, M.L. and Moore, A.W., (1996) Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research, 4, 237-285.

[23] Ceylan, H. and Bell, M. G. H. (2004) Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing. Transportation Research, 38B (4), 329–342.