Transportation Research Part C 85 (2017) 732–752

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events[☆]

Mohammad Aslani[a,*], Mohammad Saadi Mesgari[a], Marco Wiering[b]

[a] Faculty of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, Tehran, Iran
[b] Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, Groningen, The Netherlands

ARTICLE INFO

Keywords:
Adaptive traffic signal controller
Reinforcement learning
Actor-critic
Pedestrians
Traffic disruptions

ABSTRACT

The transportation demand is rapidly growing in metropolises, resulting in chronic traffic congestions in dense downtown areas. Adaptive traffic signal control as the principle part of intelligent transportation systems has a primary role to effectively reduce traffic congestion by making a real-time adaptation in response to the changing traffic network dynamics. Reinforcement learning (RL) is an effective approach in machine learning that has been applied for designing adaptive traffic signal controllers. One of the most efficient and robust type of RL algorithms are continuous state actor-critic algorithms that have the advantage of fast learning and the ability to generalize to new and unseen traffic conditions. These algorithms are utilized in this paper to design adaptive traffic signal controllers called actor-critic adaptive traffic signal controllers (A-CATs controllers).

The contribution of the present work rests on the integration of three threads: (a) showing performance comparisons of both discrete and continuous A-CATs controllers in a traffic network with recurring congestion (24-h traffic demand) in the upper downtown core of Tehran city, (b) analyzing the effects of different traffic disruptions including opportunistic pedestrians crossing, parking lane, non-recurring congestion, and different levels of sensor noise on the performance of A-CATS controllers, and (c) comparing the performance of different function approximators (tile coding and radial basis function) on the learning of A-CATs controllers. To this end, first an agent-based traffic simulation of the study area is carried out. Then six different scenarios are conducted to find the best A-CATs controller that is robust enough against different traffic disruptions. We observe that the A-CATs controller based on radial basis function networks (RBF (5)) outperforms others. This controller is benchmarked against controllers of discrete state Q-learning, Bayesian Q-learning, fixed time and actuated controllers; and the results reveal that it consistently outperforms them.

1. Introduction

The continuous population increase and subsequently the growth in social and economic activities in cities lead to the rise in demand for transportation (Bhatta, 2010). The rising demand for transportation in the metropolises has made existing traffic infrastructures incapable of handling a lot of vehicles and brings about undesired everyday traffic congestions. Traffic congestions which are produced either by the routine traffic volumes (recurring congestion) or unexpected disruptions (non-recurring congestion

[☆] This article belongs to the Virtual Special Issue on “Agent technology”.
[*] Corresponding author.
E-mail addresses: maslani@mail.kntu.ac.ir (M. Aslani), mesgari@kntu.ac.ir (M.S. Mesgari), m.a.wiering@rug.nl (M. Wiering).

http://dx.doi.org/10.1016/j.trc.2017.09.020
Received 21 December 2016; Received in revised form 22 September 2017; Accepted 22 September 2017
0968-090X/

events), mainly accidents, constructions, emergencies, break-downs, debris, or inclement weather conditions (Jeihani et al., 2015; Bifulco et al., 2016), have negative observable consequences such as long travel times, excess fuel consumption and increasing emission of local air pollutants. In order to reduce traffic congestion and its adverse negative effects, one of the most effective solutions is toward outfitting the existing infrastructure with intelligent transportation systems (ITS) which raise the capacity of existing transportation infrastructures without imposing a high cost of road construction (Chowdhury and Sadek, 2003; Bazzan and Kluegl, 2013a). ITS as systems utilizing synergistic technologies provide flexible approaches to effectively manage and control traffic. One of the significant components of ITS is adaptive traffic signal control (El-Tantawy et al., 2013; Islam and Hajbabaie, 2017).

Adaptive traffic signal control (Ma et al., 2016) is a strategy in which traffic signal timing parameters (e.g. cycle length, phase split and the duration of every controller phase) adapt based on actual traffic conditions and traffic fluctuations (e.g. number of waiting and approaching vehicles (WAVE) at the traffic signal) in order to achieve a set of specific objectives (e.g. minimizing the total number of WAVE). Adaptive traffic signal control can be modeled by self-learning in multi-agent systems (Weiss, 1999; Kluegl and Bazzan, 2012) because of the distributed and autonomous nature of traffic signal control (Adler et al., 2005; Katwijk et al., 2009). Moreover, the complexity arising in a traffic system due to the stochastic nature of the traffic patterns and complex effects of the actions performed by many other adaptive traffic signals makes it difficult to solve with preprogrammed traffic signal behaviors. Thus, a learning mechanism is necessary, such that traffic signals gradually find the global optimal solution on their own by direct interaction with the stochastic traffic environment. In this context, reinforcement learning (RL) (Sutton and Barto, 1998; van Otterlo and Wiering, 2012) as a promising method for generating, evaluating, and improving traffic signal decision making solutions is beneficial.

RL enables traffic signal controllers to learn and react flexibly to diverse traffic circumstances without the need of a predefined model of the stochastic traffic environment and also without the need of human intervention (Abdulhai and Kattan, 2003; Bazzan, 2009; El-Tantawy et al., 2014; Ozan et al., 2015; Mannion et al., 2016). Proper signal timing plans (policies) are learned through the experience of the traffic signals in their intersections (environment) rather than information retrieved from the relationship between correct input and output pairs of the traffic control system. After adaptive traffic signal controllers take actions, they receive single scalar reward signals depending on whether their actions have led them closer to realizing their objective. As a result, RL-embedded traffic signal controllers learn to obtain signal timing plans that optimize the sum of obtained rewards over time (return). By following given signal timing plans and processing the rewards, adaptive traffic signals can build estimates of the return. The function representing this estimated return is known as the value function (Sutton and Barto, 1998).

Numerous RL algorithms exist in the machine learning field. They mainly fall into one of the following three categories: 1-actor-only, 2-critic-only and 3-actor-critic, where the words actor and critic are synonyms for the policy and value function, respectively. Actor-only methods work with a parameterized family of policies. The benefit of a parameterized policy is that a spectrum of continuous actions can be generated, but the high variance in the estimation of the gradient makes learning slow, which is their weakness (Konda and Tsitsiklis, 2003). Critic-only methods such as Q-learning (Watkins and Dayan, 1992) and SARSA (Sutton and Barto, 1998) rely exclusively on value function approximation without an explicit function for the policy. Although they have a lower variance in the estimates of expected returns, to find the optimal actions in different states they need an optimization procedure in each state that makes them computationally more demanding especially if the action space is continuous. Actor-critic methods (Barto et al., 1983) are comprised of two parts, namely an actor and a critic. The critic is used to estimate the value function and the actor selects actions. The critic part evaluates the quality of the used policy and the actor uses the critic's information to update its policy parameters. Actor-critic methods have the advantages of both actor-only and critic-only methods. They are capable of producing continuous actions while the high variance in the estimation of gradients of actor-only methods is reduced by adding a critic (Grondman et al., 2012). Also, they are able to react to smoothly changing states with smoothly changing actions. In this research, actor-critic methods are employed because of their advantages over actor-only and critic-only methods.

Employing a discrete state actor-critic algorithm for traffic signal control which is naturally continuous leads to a combinatorial explosion of states and the well-known curse of dimensionality. Continuous state actor-critic algorithms use generalization that provides them with the ability of performing accurately in unseen situations. The success of continuous state actor-critic algorithms on traffic signal control hinges on effective function approximation which maps states to values via a parameterized function. Among the many function approximation schemes proposed, tile coding and radial basis functions (RBF) which strike an empirical balance between representational power and computational cost are applied in this research. Also, in order to increase the speed of learning, actor-critic methods based on eligibility traces are employed (Sutton and Barto, 1998).

Contributions of this paper. In this paper, continuous RL algorithms are applied to optimize traffic signal controllers in a traffic network modeling real variable traffic flows for 24-h on April 26, 2014, in downtown Tehran. Different actor-critic algorithms based on different types of function approximation are developed and compared on 6 different scenarios. In our traffic micro-simulations in addition to vehicles, impatient pedestrians and their interactions with vehicles are considered. Impatient pedestrians are the pedestrians who may cross junctions during red pedestrian signals if there are appropriate gaps. Impatient pedestrians may disturb vehicle movements and consequently the learning process of traffic signals by their crossing. Therefore, the effect of impatient pedestrians crossing on the traffic signals performance is examined in this research.

Since the study area is located in the administrative zone, the drivers for doing their administrative affairs usually park their vehicles beside the streets for a short time. This can in turn lead to reductions of the streets capacity. The causal effects of parked vehicles beside the streets on the learning of traffic signals are also assessed.

From a practical point of view, traffic signal's sensors can be noisy and imperfect, i.e., sensors may produce different observations from the same traffic condition. Also, reward signals as a feedback of the environment may be affected by noise. Thus, we investigate the effects of noisy states and reward signals on the performance of the learning traffic signals. Moreover, in order to make sure that