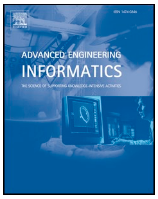
Advanced Engineering Informatics 38 (2018) 639–655

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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



Full length article

Traffic signal optimization through discrete and continuous reinforcement learning with robustness analysis in downtown Tehran

Mohammad Aslani[a,*], Stefan Seipel[a,b], Mohammad Saadi Mesgari[c], Marco Wiering[d]

[a] *Department of Industrial Development, IT and Land Management, University of Gavle, Gavle, Sweden*
[b] *Division of Visual Information and Interaction, Department of Information Technology, Uppsala University, Uppsala, Sweden*
[c] *Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran*
[d] *Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, Groningen, The Netherlands*

ARTICLE INFO

Keywords:
Reinforcement learning
System disturbances
Traffic signal control
Microscopic traffic simulation

ABSTRACT

Traffic signal control plays a pivotal role in reducing traffic congestion. Traffic signals cannot be adequately controlled with conventional methods due to the high variations and complexity in traffic environments. In recent years, reinforcement learning (RL) has shown great potential for traffic signal control because of its high adaptability, flexibility, and scalability. However, designing RL-embedded traffic signal controllers (RLTSCs) for traffic systems with a high degree of realism is faced with several challenges, among others system disturbances and large state-action spaces are considered in this research.

The contribution of the present work is founded on three features: (a) evaluating the robustness of different RLTSCs against system disturbances including incidents, jaywalking, and sensor noise, (b) handling a high-dimensional state-action space by both employing different continuous state RL algorithms and reducing the state-action space in order to improve the performance and learning speed of the system, and (c) presenting a detailed empirical study of traffic signals control of downtown Tehran through seven RL algorithms: discrete state Q-learning(λ), SARSA(λ), actor-critic(λ), continuous state Q-learning(λ), SARSA(λ), actor-critic(λ), and residual actor-critic(λ).

In this research, first a real-world microscopic traffic simulation of downtown Tehran is carried out, then four experiments are performed in order to find the best RLTSC with convincing robustness and strong performance. The results reveal that the RLTSC based on continuous state actor-critic(λ) has the best performance. In addition, it is found that the best RLTSC leads to saving average travel time by 22% (at the presence of high system disturbances) when it is compared with an optimized fixed-time controller.

1. Introduction

Traffic control is of a vital importance in densely populated cities. The ineffective control of traffic can cause significant costs for drivers owing to the increased wasted time, negative effects on the environment due to vehicle emissions and detrimental impacts on the economy due to increased fuel consumption [1]. The main components of a traffic control system are traffic signal control, ramp-metering, variable speed limit enforcement, and dynamic route guidance [2]. Within such a context, scheduling of traffic signals is a critical traffic control challenge. Inefficiency in traffic signal timing attributed to the inability of adapting to prevailing traffic conditions leads to different small congested areas that can in turn cause larger traffic jams [3]. In recent years, computational intelligence techniques such as fuzzy logic [4–6], neural networks [7,8], and reinforcement learning (RL) [9–11] have shown their potential for designing adaptive traffic signal controllers. In this paper, RL [12] is applied because of its online learning ability to gradually improve its performance, its adaptability to different traffic states as well as its ability to work without knowing an explicit model of the stochastic traffic environment. An RL-embedded traffic signal controller (RLTSC) has the capability to learn through experience by dynamically interacting with the traffic environment in order to reach its objective. Each RLTSC examines different green time durations (actions) in different traffic situations (states) and determines the best sequence of them based on the received scalar reward signals (feedback) which indicate the quality of the selected action. Each RLTSC which controls one intersection learns over time to obtain a signal timing plan that optimizes the sum of rewards in the future (return).

* Corresponding author at: Department of Industrial Development, IT and Land Management, University of Gavle, Gavle, Sweden.
E-mail addresses: mohammad.aslani@hig.se, moh.aslani@gmail.com (M. Aslani), Stefan.Seipel@hig.se (S. Seipel), mesgari@kntu.ac.ir (M.S. Mesgari), m.a.wiering@rug.nl (M. Wiering).

https://doi.org/10.1016/j.aei.2018.08.002
Received 21 June 2017; Received in revised form 15 July 2018; Accepted 14 August 2018

RL algorithms can be either model-based [13] or model-free [9,14,15]. Model-based approaches need to initially employ and/or learn the environmental model (i.e. traffic system) in order to compute the optimal policy (signal timing plan). On the other hand, a model-free approach (e.g. SARSA, Q-learning, and actor-critic) does not rely on the estimation of the environmental model; instead, it progressively acquires the optimal policy by interacting with the environment and getting experience [16]. Both approaches have their strengths and weaknesses regarding their convergence guarantees, convergence speed, and ability to plan [12]. However, in the model-based approaches, obtaining an accurate model of the environment can be challenging and a slight bias in the model may lead to a strong bias in the policy. Thus, we employ model-free approaches, including Q-learning, SARSA, and actor-critic, which also makes the research more appealing for field deployment.

Conventional model-free RL algorithms require storing distinct estimations of each state-action value (for SARSA and Q-learning algorithms) or each state value (for the critic part of actor-critic algorithms) in lookup tables. Although they are computationally less demanding, their learning process can be slow when the state-action space is large. The reason is that the agent needs to experience all possible states. In this context, there are a couple of solutions for coping with a large state-action space. A first step to the solution is to employ continuous-state RL. In this version of RL, the knowledge gained from observations in each state (traffic condition) is applied to similar states by means of function approximators [17]. The application of previously acquired knowledge to unseen states generally leads to faster convergence. To handle the complexity of the state space, the state space is first mapped onto a feature space by using a feature function. Then the values of states are approximated in the feature space, instead of the original state space. Finding good function approximators of appropriate complexity is the key to the success of continuous state RL. The second solution is to reduce the state-action space by efficiently removing unnecessary and less effective state variables and actions in order to decrease the number of interactions needed to learn the optimal timing plan. An ineffective reduction of the state-action space may lead to undesired results because it cannot provide the agent with the required information that might be useful in decision making. Thus, it should be investigated which state variables and actions based on the conditions of the study area should be included in the state-action space so that optimal signal timing plans can be learned successfully. However, even if some trivial state variables are eliminated, the state-action space can still be too large for successfully being handled in discrete state RL. The third solution which is the combination of solutions 1 and 2 uses both state-action space reduction and continuous state RL. The first contribution of the present paper is to adopt all these three solutions and compare their results.

Regarding function approximation in continuous state RL, there are several different types of function approximators that are categorized into linear and non-linear function approximators. The values of states are represented by a weighted linear sum of a set of non-linear extracted features in linear function approximators or are sometimes represented by non-linear approximators such as neural networks. While non-linear function approximators may approximate an unknown function with better accuracy, linear function approximators are better to understand, simpler to implement, and faster to compute [18]. Also, linear function approximators are able to estimate non-linear value functions due to the employed basis function [12]. Because of the advantages of linear function approximators, linear function approximators are employed in this paper. A popular method to extract features in a linear function approximator is tile coding that splits the state space into separate tiles and assigns one feature to each tile [19]. Striking an empirical balance between representational power and computational cost is one of the most important advantages of tile coding that makes it suitable to be employed in this research.

In this article, a detailed empirical study of traffic signals control in downtown Tehran through seven discrete and continuous state RL algorithms, namely discrete state Q-learning(λ), discrete state SARSA(λ), discrete state actor-critic(λ), continuous state Q-learning(λ), continuous state SARSA(λ), continuous state actor-critic(λ) [12,18], and continuous state residual actor-critic(λ) [20] is conducted. The traffic congestion in downtown Tehran is very heavy. Its traffic control systems are old-fashioned and most traffic signals are still manually controlled by police officers that can lead to inefficient traffic control.

One of the most significant challenges in designing the RLTSC, is the robustness of them against different system disturbances that almost always happen in real-world applications. System disturbances, based on their intensity level, can disturb either the normal performance or even the convergence of the system. We consider three different kinds of system disturbances: jaywalking, incidents, and sensor noise. In the study area, observation surveys illustrated that many pedestrians prefer making the illegal crossing to waiting for the green pedestrian signal. Jaywalking which is performed by impatient pedestrians during the red pedestrians' signal increase the stochasticity of the traffic environment and disturb the performance of the RLTSCs. Incidents as a nonrecurring traffic congestion cause [21] that disturb the traffic flow, make unexpected delays in the movements of the vehicles and consequently make the traffic environment more nonstationary for the RLTSCs. Moreover, due to the obsolescence of the traffic control infrastructures, traffic signal sensors can be noisy and imperfect. In fact, the RLTSCs' observations of the state as well as the reward signal are noisy. That is, the observed number of vehicles waiting on the approaching streets are different than their true values. Sensor noise, unlike the first two factors that have an indirect effect, directly disrupts the RLTSCs' performance. This represents a difficult class of learning problem owing to the stochastic nature of the traffic environment together with high-order dynamics (i.e. variable traffic flows) and sensor noise. The RLTSCs should be able to autonomously respond to a changing environment with stochasticity and random shocks. In the literature, different RLTSCs have been investigated, but without thoroughly examining the robustness of them against system disturbances. Thus, another contribution of this paper is the comprehensive investigation of the RLTSCs' robustness to the system disturbances. In this context, a real-world microscopic traffic simulation of the upper downtown core of Tehran city is carried out. In this simulation, a three-dimensional traffic network (i.e. slope of the streets are considered) with real changing traffic flows over 6 h in the morning (6:00 am to 12:00 pm) is considered.

Outline of this paper. The remaining part of this paper is organized as follows: Section 2 reviews the related work and summarizes the gaps in the existing literature. Section 3 describes the operation of discrete and continuous state RL. Section 4 demonstrates the traffic network and microscopic traffic simulation of the central area of Tehran. Section 5 technically explains the design of the RLTSCs and the way they work. System disturbances, namely incidents, jaywalking, and sensor noise are explained in Section 6. Section 7 describes a series of experiments and their results and Section 8 provides a discussion of our findings. Finally, Section 9 concludes the paper and proposes some directions for future work.

2. Related work

Traffic signal control has been one of the major challenges for controlling traffic congestion in urban areas. Different systems and methods based on complex mathematical models have been proposed in order to optimize traffic signal parameters in response to diverse traffic situations. The Split Cycle Offset Optimization Technique (SCOOT) [22] and the Sydney Coordinated Adaptive Traffic System (SCATS) [23] are two successful commercial systems that have been installed in more than one hundred cities worldwide. Although these systems have alleviated traffic congestion somewhat, they may not be efficient in handling the traffic networks without well-defined traffic flow patterns (e.g. morning and afternoon peak hours). Also, they