Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 109C (2017) 513–520

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Ambient Systems, Networks and Technologies (ANT 2017)

Adaptive Traffic Signal Control : Exploring Reward Definition For Reinforcement Learning

Saad Touhbi[a,b,c,*], Mohamed Ait Babram[a,b], Tri Nguyen-Huu[b], Nicolas Marilleau[b], Moulay L. Hbid[a,b], Christophe Cambier[b,c], Serge Stinckwich[b,c,d]

[a] *LMDP/IRD Cadi Ayyad University, Marrakech, Morocco*
[b] *IRD, UMI 209, UMMISCO, IRD France Nord, F-93143, Bondy, France*
[c] *Sorbonne Universites, Univ. Paris 06, UMI 209, UMMISCO, F-75005,Paris,France*
[d] *Universite de Caen Basse-Normandie, Caen, France*

Abstract

As mobility grow in urban cities, traffic congestion become more frequent and troublesome. Traffic signal is one way to decrease traffic congestion in urban areas but needs to be adjusted in order to take into account the stochasticity of traffic. Reinforcement learning (RL) has been the object of investigation of many recent papers as a promising approach to control such a stochastic environment. The goal of this paper is to analyze the feasibility of RL, particularly the use of Q-learning algorithm for adaptive traffic signal control in different traffic dynamics. A RL control was developed for an isolated multi-phase intersection using a microscopic traffic simulator known as Paramics. The novelty of this work consists of its methodology which uses a new generalized state space with different known reward definitions. The results of this study demonstrate the advantage of using RL over fixed signal plan, and yet exhibit different outcomes depending on the reward definitions and different traffic dynamics being considered.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Reinforcement learning; traffic optimization; traffic light control; Q-learning; urban mobility

1. Introduction

Urban cities are usually covered by a complex traffic network responsible for supporting the daily demands in the area. Unfortunately, the traffic demands are high, dynamic and in constant increase. Infrastructure improvement has been the primary method to serve these demands. However, this is not always possible due to constraints such as financial resources and space. This has led to options considering the improvement of the existing infrastructure, optimizing the utilization of the available infrastructure and lowering the costs of travel time through intelligent transportation systems (ITS). Adaptive traffic signal control is most used since the early seventies[1] and has shown to be

* Corresponding author.
E-mail address: saad.touhbi@edu.uca.ma

Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.05.327

a suited approach to alleviate traffic congestion as opposed to pre-timed and actuated control systems for signalized intersections. Many methods have been developed and investigated in the literature (SCOOT[2] and SCATS[3], PRODYN[4], OPAC[5], UTOPIA[6], RHODES[7]) but these methods require a pre-specified model of the environnement. But due to the stochastic nature of traffic, An approach that can adapt to the changes in traffic and does not need a specified model for a certain environnement would be more simple for the control process. Reinforcement learning[8] on the other hand has the ability to adapt and self-learn from past experiences. Therefore it has more potential to improve service over time through continuous interaction with the environnement. In this paper, we will investigate the use of reinforcement learning, in particular Q-learning on trafic signal control. The study will include previous work with a new representation of state space based on queuing levels. We will also look into the effect of different reward definitions on traffic patterns.

2. Related Work

Traffic signal systems are most used in intersections, especially in urban areas. Generally, traffic signals go through repeated *"cycles"*, each cycle corresponding to one complete rotation through all of the indications provided at the intersection. Each cycle consists of a sequence of N phases. A *"phase"* is a group of traffic flow movement through the intersection (e.g. North to South and South to North); a typical phase has a time interval of passage defined as *"green"* time, followed by a duration of *"yellow"* and then *"red"* (red can signify the passage to the next phase or a halt time for the intersection to be empty; in this case we call it *"All red"*). The duration is determined for each phase either as a fixed plan using the commonly used Webster method[9], or using adaptive allocation of green time for each phase and variable sequences of phases depending on traffic dynamics[2,3,4,5,6,7]. Application of RL to adaptive traffic light control has been proven to be efficient in many papers[10]. SARSA for traffic light control was introduced by Thorpe[11], Abdulhai et al.[12] introduced Q-learning for an isolated intersection. Bingham[13] introduced a neuro-fuzzy traffic signal controller that uses RL for learning the neural network. Olivera et al.[14], proposed an RL method with context detection to solve signal control optimization. These previous studies[10,11,12,13,14] are designed to solve fixed phasing sequence for signal control and have been used on hypothetical intersections to show their feasibility. However, El-Tantawy et al.[15] used several RL algorithms for a comprehensive analysis on the effect of state space, action space, action selection and reward definition on patterns of traffic on a real world intersection, and proved that variable phasing can be more efficient especially in high and dynamic traffic patterns. Brys et al.[16] used variable phasing and alternative reward definition to solve the traffic signal control on hypothetical road network. In this paper we extend the previous work done by El-tantawy et al.[15] by proposing at first a new state space definition that could provide a more standard representation of the state of the intersection. Additionally, this paper provides an analytical study of the effect of reward definition and traffic patterns on the performance of RL in an isolated intersection, which could lead to an in-depth study on traffic variability and its effect on the RL controlled traffic intersections.

3. Proposed System

For our study, we will use the intersection in Downtown Toronto (Front and Bay Street) used by El-Tantawy et al.[15] with two different traffic volumes. Our performance metrics are the average delay experienced by each vehicle, the rate of vehicles passing through the intersection, i.e. throughput and the average queue length.

3.1. Q-learning

Q-learning is a RL technique that is used to estimate an optimal action-selection policy for a given finite state space S and a set of actions A. Q-learning is an off policy (model free) algorithm since the agent attempts to improve a policy with no required knowledge on the system. Performing an action in a specific state provides a reward; the goal of the algorithm is to maximize the long-term reward. This algorithm learns by updating an action-function $Q : S \times A \mapsto \mathbb{R}$ that gives the expected utility of taking an action $a \in A$ on a given state $s \in S$ which eventually return after enough