Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 130 (2018) 26–33

Procedia
Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Ambient Systems, Networks and Technologies, ANT-2018 and
the 8th International Conference on Sustainable Energy Information Technology,
SEIT 2018, 8-11 May, 2018, Porto, Portugal

Evaluating reinforcement learning state representations for adaptive traffic signal control

Wade Genders[a], Saiedeh Razavi[a]

[a]*Civil Engineering, McMaster University, 1280 Main St. West, Hamilton, L8S 4L8, Canada*

Abstract

Reinforcement learning has shown potential for developing effective adaptive traffic signal controllers to reduce traffic congestion and improve mobility. Despite many successful research studies, few of these ideas have been implemented in practice. There remains uncertainty about what the requirements are in terms of data and sensors to actualize reinforcement learning traffic signal control. We seek to understand the data requirements and the performance differences in different state representations for reinforcement learning traffic signal control. We model three state representations, from low to high-resolution, and compare their performance using the asynchronous advantage actor-critic algorithm with neural network function approximation in simulation. Results show that low-resolution state representations (e.g., occupancy and average speed) perform almost identically to high-resolution state representations (e.g., individual vehicle position and speed). These results indicate implementing reinforcement learning traffic signal controllers may be possible with conventional sensors, such as loop detectors, and do not require sophisticated sensors, such as cameras or radar.

Keywords: Adaptive Traffic Signal Control; Reinforcement Learning; Artificial Neural Networks; Intelligent Transportation Systems; Traffic Simulation;

1. Introduction

Vehicle congestion is a major problem in cities across the world. Developing additional infrastructure is expensive and a protracted process which can exacerbate the problem until completed. Instead of adding more infrastructure, another solution is to optimize currently available infrastructure. Intersection traffic signal controllers (TSC) are ubiquitous in modern road infrastructure and their functionality greatly impacts all users. Many research studies have proposed improvements to TSC, broadly in an attempt to make them adaptive to current traffic conditions. Reinforcement learning has been shown to be effective in developing adaptive TSC with many research studies detailing

* Corresponding author: Wade Genders
E-mail address: genderwt@mcmaster.ca

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2018.04.008

promising results. Despite the encouraging research, few reinforcement learning adaptive TSC have been deployed in the field. One inhibiting factor is the resources required; to observe the traffic state, reinforcement learning TSC often require high-resolution data beyond the detection capability of traditional sensors (i.e., loop detectors). This research focuses on the potential state definitions of reinforcement learning TSC and ascertaining the performance differences between them. We seek to answer, can a reinforcement learning TSC function using low-resolution data from traditional sensors such loop detectors? Or is high-resolution data from sophisticated sensors (e.g., cameras, radar) required? Answering this question will help individuals interested in deploying reinforcement learning TSC in the field, as they will be aware of the requirements and potential outcomes. We use the traffic microsimulator SUMO[1] and the asynchronous advantage actor-critic (A3C) algorithm[2] to train and evaluate multiple adaptive TSC with different resolution state representations.

2. Literature Review

Many research studies have recognized and displayed reinforcement learning's capability for providing a solution to TSC. Early research provided proof-of-concept for reinforcement learning in TSC[3,4,5,6]. Later research applied reinforcement learning methods to more realistic and complex traffic models[7,8,9,10,11,12]. Developments in machine learning have yielded deep reinforcement learning techniques[2,13,14] which have subsequently been applied for TSC[15,16,17,18,19].

Considering the aforementioned research and the extensive reinforcement learning TSC reviews[20,21,22], we identify numerous possible state representations: vehicle density, flow, queue, location, speed along with the current traffic phase, cycle length and red time. These state representations form a resolution spectrum of the current traffic state, from coarse (e.g., flow) to fine (e.g., individual vehicle position and speed). We consider state representation across the resolution spectrum, requiring different sensors, and compare their performance. The results can guide individuals interested in practical implementation.

3. Model

3.1. Reinforcement Learning

Reinforcement learning is a type of machine learning for solving sequential decision-making problems[23]. A reinforcement learning agent learns a policy $\pi(s) = a$, mapping from states s to actions a, to achieve a goal in an environment under uncertainty. Through repeated environment interactions, a reinforcement learning agent strives to develop an optimal policy π^*, which maximizes the sum of future discounted ($\gamma \in (0, 1]$) rewards, defined as the return G_t in Equation 1:

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{1}$$

The agent interacts with the environment in repeating sequences of, at time t, observing the environment state s_t, taking action a_t, receiving reward r_t and entering a new state s_{t+1}. Over time, the agent learns what actions in what states maximize long-term reward, also known as value. Rewards quantitatively represent how successful the agent's policy is achieving its mandated goal.

The A3C algorithm is used to develop parameterized θ policy $\pi(a|s;\theta)$ (Equation 2) and value $V^{\pi}(s;\theta)$ functions (Equation 3). The agent develops a value function (critic), which estimates the expected return from a given state, which is used to improve the policy (actor).

$$\pi(a|s;\theta) = \Pr[a_t = a|s_t = s;\theta] \tag{2}$$

$$V^{\pi}(s;\theta) = \mathbb{E}[G_t|s_t = s;\theta] \tag{3}$$

The parameters are used for neural network function approximation, defining the weights between neurons.