Engineering Applications of Artificial Intelligence 29 (2014) 134–151

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework

Mohamed A. Khamis [a,*], Walid Gomaa [a,b]

[a] Department of Computer Science and Engineering, Egypt-Japan University of Science and Technology (E-JUST), P.O. Box 179, New Borg El-Arab City, Postal Code 21934 Alexandria, Egypt
[b] Alexandria University, Alexandria 21544, Egypt

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
4 December 2013
Accepted 9 January 2014
Available online 1 February 2014

Keywords:
Adaptive optimization
Multi-objective optimization
Reinforcement learning
Exploration
Traffic signal control
Cooperative multi-agent system

ABSTRACT

In this paper, we focus on computing a consistent traffic signal configuration at each junction that optimizes multiple performance indices, i.e., multi-objective traffic signal control. The multi-objective function includes minimizing trip waiting time, total trip time, and junction waiting time. Moreover, the multi-objective function includes maximizing flow rate, satisfying green waves for platoons traveling in main roads, avoiding accidents especially in residential areas, and forcing vehicles to move within moderate speed range of minimum fuel consumption. In particular, we formulate our multi-objective traffic signal control as a multi-agent system (MAS). Traffic signal controllers have a distributed nature in which each traffic signal agent acts individually and possibly cooperatively in a MAS. In addition, agents act autonomously according to the current traffic situation without any human intervention. Thus, we develop a multi-agent multi-objective reinforcement learning (RL) traffic signal control framework that simulates the driver's behavior (acceleration/deceleration) continuously in space and time dimensions. The proposed framework is based on a multi-objective sequential decision making process whose parameters are estimated based on the Bayesian interpretation of probability. Using this interpretation together with a novel adaptive cooperative exploration technique, the proposed traffic signal controller can make real-time adaptation in the sense that it responds effectively to the changing road dynamics. These road dynamics are simulated by the Green Light District (GLD) vehicle traffic simulator that is the testbed of our traffic signal control. We have implemented the Intelligent Driver Model (IDM) acceleration model in the GLD traffic simulator. The change in road conditions is modeled by varying the traffic demand probability distribution and adapting the IDM parameters to the adverse weather conditions. Under the congested and free traffic situations, the proposed multi-objective controller significantly outperforms the underlying single objective controller which only minimizes the trip waiting time (i.e., the total waiting time in the whole vehicle trip rather than at a specific junction). For instance, the average trip and waiting times are $\simeq 8$ and 6 times lower respectively when using the multi-objective controller.

1. Introduction

In this paper, we focus on computing a consistent traffic signal configuration at each junction that optimizes multiple performance indices (i.e., multi-objective traffic signal control). Traffic signal control can be viewed as a multi-objective optimization problem. The multi-objective function can have a global objective for the entire road network or there may be different objectives for the different parts of the road network (e.g., maximize safety especially in residential and schools areas), or even different times of the day for the same part of the road network.

Construction of a new infrastructure is expensive, thus the generally acceptable solution is to improve the utilization of the existing resources by moving towards *Intelligent Transportation Systems* (*ITS*) for traffic management and control. Traffic control is a set of methods that are used to enhance the traffic network performance by, for example, controlling the traffic flow to minimize congestion, waiting times, fuel consumption and avoid accidents. Traffic control generally includes the following components; controlling the traffic signals in urban areas, ramp-metering in highways, enforcing variable speed limits (according to vehicles types), supporting the drivers with route guidance based on the up-to-date traffic status using some kind of navigation systems (e.g., GPS), enforcing overtaking rules, and using driver-assistance systems (e.g., adaptive cruise control). In this paper, we particularly focus on controlling traffic signals in urban areas.

Another two important components of the ITS are traffic modeling and traffic simulation. Traffic modeling is the formulation of rigorous mathematical models that represent the various dynamics of the traffic system. This includes drivers' behavior in acceleration, deceleration, lane changing, phenomena such as

* Corresponding author. Tel.: +20 3 309 4075; Mobile: +20 100 638 2428; fax: +20 3 459 9520.
E-mail addresses: mohamed.khamis@ejust.edu.eg (M.A. Khamis), walid.gomaa@ejust.edu.eg (W. Gomaa).

0952-1976/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2014.01.007

rubbernecking, and behavior change under different weather conditions. Traffic simulation is the virtual emulation of the traffic system on digital computers. Traffic simulators are used for experimentation and validation of the underlying traffic models and traffic control mechanisms.

Intelligent traffic control has many challenges that include the continuing increase in the number of vehicles (it is expected that 70% of the people worldwide will live in urban areas by 2050, Pizam, 1999), the high dynamics and non-stationarity of the traffic network, and the nonlinear behavior of the different components of the control system.

Nowadays, the different types of transportation means (specifically vehicles in urban areas) have major problems that governments are facing in both developing and developed countries. Traffic of vehicles in urban areas, specifically, has many problems that include increase of traffic congestion, psychological stress of drivers that affects their behavior leading to a high rate of accidents, considerable time losses, and a high rate of vehicle emissions which severely affects the environment. Those problems have a considerable negative effect on the country economy. Thus, in this paper, the proposed traffic signal controller tackles most of those problems (e.g., minimizes the waiting time of vehicles) as will be shown by the performance evaluation in Section 7.

In 2010, traffic costs (based on time loss and fuel consumption) about $115 billion in the US based on 439 urban areas (Schrank et al., 2011). In the same year, 32,885 people died in accidents in the US (U.S. Department of Transportation, 2012). In Egypt, traffic problems are responsible for more than 25,000 accidents in 2010 with more than 6000 deaths per year (CAPMAS). Deaths per million driving kilometers in Egypt is about 34 times greater than in the developed countries (Abbas, 2004). This value is about 3 times greater than countries in the Middle East region (Abbas, 2004). The authors expect that this value is much worse in 2011–2013 due to the political upheaval in Egypt.

Recently, some computer science tools and technologies have been used to address the traffic signal control complexities. Among these is the MAS framework whose characteristics are similar in nature to the traffic problem (Shoham and Leyton-Brown, 2010; De-Oliveira and Camponogara, 2010). Such characteristics include distributivity, autonomy, intelligibility, on-line learnability, and scalability. In particular, the formulation of the traffic signal control problem as a multi-agent reinforcement learning (MARL) configuration is very promising (as proposed in Bazzan, 2009).

In the current paper, we adopt a MARL framework in a cooperation-based configuration to comply with the distributed nature and complexity of the problem. Our work is a significant extension of the framework developed by Wiering (2000) and Wiering et al. (2004). Wiering's controller, namely TC-1, represents a *pioneering* step in the use of real-time reinforcement learning framework in modeling traffic signal control. TC-1 outperforms traditional controllers (e.g., random, fixed time, longest queue, most cars). Moreover, TC-1 has proved its effectiveness and efficiency when being applied to large scale traffic networks. In contrast, other controllers based on reinforcement learning, e.g., Thorpe and Anderson (1996) and Abdulhai et al. (2003) suffer from exponential state-spaces when applied to large scale traffic networks. In addition, many latter researchers, e.g., Houli et al. (2010), Kuyer et al. (2008), Schouten and Steingröver (2007), Iša et al. (2006), and Steingröver et al. (2005), use TC-1 as a *benchmark* for performance evaluation. Each of these controllers contribute to TC-1 from a different prospective. For instance, in Schouten and Steingröver (2007), the authors overcome the *partial observability* of the traffic state-space, while we assume that the state-space is *fully-observable*, i.e., the agent can perfectly sense its environment.

Nevertheless, as will be explained latter, we tackle some problems in which TC-1 fails to adapt with. This includes: (1) stable adaptation to the limited-time congestion periods (using Bayesian probability interpretation), (2) advanced reward formulation to adapt with the continuous-time continuous-space simulation platform, and (3) using a multi-objective reward formulation in an additive manner to optimize multiple performance indices.

Moreover, we evaluate the performance of our proposed controller in comparison with two adaptive control strategies which are also based on AI methods: Self-Organizing Traffic Lights (SOTL) (Cools et al., 2008) (that outperforms a traditional green wave controller) and a Genetic Algorithm (GA) (Wiering et al., 2004).

Particularly, our objective in this paper is to develop a traffic control framework with the following characteristics: (1) inherently distributed through the use of a vehicle-based *multi-agent system*; there are two types of agents: *traffic junction agents* (active computing agents) which are responsible for the decision making process (i.e., deciding on the proper traffic signal configuration) according to the information collected from the vehicle agents; *vehicle agents* (passive agents) which support the decision making process by communicating the necessary information to the junction agents, (2) online sequential decision making framework where decisions are taken in real-time for signal splitting based on multiple optimization criteria; the core of the applied mechanism is based on Dynamic Programming (DP) which is very-well suited for sequential decision making tasks; the real-time optimization and decision making is done incrementally by integrating the online learning with DP through the use of reinforcement learning, (3) effectively and efficiently handle the inherent complexity of the problem, the uncertainties involved, the incompleteness of information, the absence of a rigorous modeling of the traffic volume and the general dynamics: through the use of stochastic and statistical tools to predict the unknown parameters and provide an up-to-date model of the current traffic conditions, (4) adaptive system in the sense that it responds effectively to the road dynamics (variations in traffic demand, changing weather conditions, etc.): through the use of a Bayesian approach for estimating the parameters of the underlying Markov Decision Process (MDP) and the use of an adaptive cooperative hybrid exploration technique, and (5) higher confidence in the validity of the proposed traffic signal controller: through the use of a more realistic simulator as a testbed that is achieved by implementing the IDM acceleration model (Treiber et al., 2000) in the GLD vehicle traffic simulator (Wiering et al., 2004); moving from the unrealistic discrete-time discrete-space simulation platform to a continuous-time continuous-space one.

The discrete-time discrete-space simulation platform was unrealistic in the sense that the first waiting vehicle jumps once the traffic signal turns green. Now, by applying the more realistic IDM acceleration model, the vehicle takes the normal time to decelerate when a traffic signal turns red and accelerates back again to cross the junction when the signal turns green. This behavior, on the other side, causes some kind of sign oscillation when being applied on the underlying RL model as will be shown later in Section 4 (which we called the *Zeno phenomenon*[1]) which results from the very slow acceleration of back vehicles when the traffic signal is just turning green.

Preliminary results of this work have been published in Khamis et al. (2012a,b) and Khamis and Gomaa (2012). In this paper, we provide a more detailed description and improvements on the multi-objective function. Such improvements boost the performance of the multi-objective controller, particularly when being compared to the

[1] A Zeno phenomenon occurs due to the infinitesimal motion of a particle continuously within the same state.