Transportation Research Part C 88 (2018) 257–271

Contents lists available at ScienceDirect

Transportation Research Part C

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

Analysing the impact of travel information for minimising the regret of route choice☆

Gabriel de O. Ramos*, Ana L.C. Bazzan, Bruno C. da Silva

Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

A R T I C L E I N F O

Keywords:
Route choice
Regret minimisation
Multiagent reinforcement learning
Regret estimation
Mobile navigation app
Travel information

A B S T R A C T

In the route choice problem, self-interested drivers aim at choosing routes that minimise travel costs between their origins and destinations. We model this problem as a multiagent reinforcement learning scenario. Here, since agents must adapt to each others' decisions, the minimisation goal is seen as a moving target. Regret is a well-known performance measure in such settings, and considers how much worse an agent performs compared to the best fixed action in hindsight. In general, regret cannot be computed (and used) by agents because its calculation requires observing the costs of all available routes (including non-taken ones). In contrast to previous works, here we show how agents can compute regret by building upon their experience and via information provided by a mobile (Waze-like) navigation app. Specifically, we compute the regret of each action as a linear combination of local (experience-based) and global (app-based) information. We refer to such a measure as the *action regret*, which can be used by the agents as reinforcement signal. Under these conditions, agents are able to minimise their external regret even when the cost of routes is *not* known in advance. Based on experimental evaluation in several abstract road networks, we show that the system converges to approximate User Equilibria.

1. Introduction

The *route choice problem* concerns how rational drivers[1] behave when choosing routes between their origins and destinations to minimise their travel costs. In order to accomplish this goal, drivers must adapt their choices to account for changing traffic conditions. Such scenarios are naturally modelled as multiagent systems. Multiagent reinforcement learning (RL) captures the idea of self-interested agents interacting in a shared environment to improve their outcomes. In the basic, single-agent RL setting, the agent must learn by trial-and-error how to behave in an environment in order to maximise its utility. However, in multiagent RL settings, multiple agents share a common environment, and thus must adapt their behaviour to each other. In other words, the learning objective of each agent becomes a moving target.

An interesting class of multiagent RL techniques comprises the regret minimisation approaches. Different notions of regret are considered in the literature (Blum and Mansour, 2007). The most common one is that of *external regret*, which measures how much worse an agent performs on average in comparison to the best fixed action in hindsight. In this sense, regret minimisation can be seen as an inherent definition on how rational agents behave over time.

The use of regret in the context of route choice and correlated problems has been widely explored in the literature (Cesa-Bianchi

☆ This article belongs to the Virtual Special Issue on "Agents in Traffic and Transportation".

* Corresponding author.
E-mail addresses: goramos@inf.ufrgs.br (G.de O. Ramos), bazzan@inf.ufrgs.br (A.L.C. Bazzan), bsilva@inf.ufrgs.br (B.C. da Silva).

[1] Henceforth, we use the terms *agent* and *driver* equivalently.

https://doi.org/10.1016/j.trc.2017.11.011
Received 31 January 2017; Received in revised form 1 November 2017; Accepted 10 November 2017
0968-090X/

and Lugosi, 2006). In the transportation literature, regret has been employed to develop discrete choice models that predict travellers' behaviour (Chorus et al., 2008; Chorus, 2012). However, as opposed to our approach, such models focus on the traffic manager (centralised) perspective, assuming full knowledge and ignoring agents' adaptation. Within RL, regret has been mainly employed as a performance measure (Bowling, 2005; Banerjee and Peng, 2005; Prabuchandran et al., 2016). In contrast to such approaches, here we use regret to *guide* the learning process. A few existing techniques indeed use regret as reinforcement signal (Zinkevich et al., 2008; Waugh et al., 2015), but they assume regret is known a priori by the agents. However, we highlight that calculating regret requires complete knowledge about the environment (i.e., the reward associated with every possible action along time). On the one hand, such knowledge may be obtained using on-line services (Vasserman et al., 2015; Hasan et al., 2016), which provide travel information to end-users through mobile navigation apps. On the other hand, computing regret in the absence of global information is more challenging (Stone and Veloso, 2000), especially in highly competitive scenarios like traffic. We combine these two fronts and investigate how agents can estimate their regret using both local (experience-based) and global (app-based) information. Previously, we investigated a similar direction (Ramos and Bazzan, 2016) and provided formal performance guarantees (Ramos et al., 2017). However, agents' regret was computed using only local information.

In this paper, we develop a regret minimisation algorithm for handling the route choice problem. We propose a method for agents to estimate their regret using both local information (an internal history of observed rewards) and global information (travel times provided by a mobile navigation app). Apart from agent regret, we also consider the *action regret*, which measures the average amount lost by an agent up to a given time for taking a specific action rather than the best one. The action regret can thus be used for updating agents' policies. The expected system's outcome corresponds to the so-called User Equilibrium (UE) (Wardrop, 1952), i.e., an equilibrium point in the space of policies in which no driver benefits from deviating from its policy. To the best of our knowledge, this is the first approach in which agents compute their regret by combining local and global information and without assuming that the cost of all possible actions, in all situations, is known in advance.

The main hypotheses of our work are that: (i) learning to minimise regret improves drivers' performance, (ii) using app-based information reduces regret, and (iii) the system converges to an approximate UE. In the present setting, *learning* means finding the best route to take. Recall that this objective is a moving target because there are many agents interacting within the same environment. In this respect, convergence to a solution means a point at which agents keep *exploiting* their knowledge most of the time and the system is somewhat *stable* (i.e., agents only observe small fluctuations in their costs). Our key contribution is to show that when the agents are using our approach such a stable point is close to the UE. In particular, the contributions of this work are:

- We define a (mobile) navigation entity (*app*, henceforth) that provides travel information[2] to the agents. Information here is simply the average travel times of the routes used by the agents. Such information is useful for the agents to estimate their regret.
- We propose an action-based measure of regret, the *action regret*, which can be used as reinforcement signal in the RL process.
- We introduce a method for agents to *estimate* their action regret using a linear combination of their experience (rewards received in previous episodes) and information provided by the app. We show that such estimates can be used to improve the learning process.
- We develop an RL algorithmic solution that updates an agent's policy using *action regret as reinforcement signal*. Consequently, the agents learn to choose actions that minimise their external regret.

This paper is organised as follows. Section 2 provides a background on the topics related to this work. Section 3 presents the proposed methods. Our approach is experimentally evaluated in Section 4. Concluding remarks are presented in Section 5.

2. Background

In this section, we review the literature on route choice (Section 2.1), reinforcement learning (Section 2.2), regret minimisation (Section 2.3) and routing with non-local information (Section 2.4).

2.1. The route choice problem

The route choice problem concerns how drivers behave when choosing routes between their origins and destinations (OD pair, henceforth). In this section, we introduce the basic concepts related to route choice.

A road network can be represented as a directed graph $G = (N,L)$, where the set of nodes N represent intersections and the set of links L represent roads between intersections. The demand for trips generates a flow of vehicles on the links, where f_l is the flow on link l. A trip is performed by means of a route R, which is a sequence of links connecting an OD pair. Each link $l \in L$ has a cost $c_l : f_l \to \mathbb{R}^+$ associated with it. The cost of a route R is $C_R = \sum_{l \in R} c_l$. Such costs are typically modelled using a volume-delay function (VDF). A well-known VDF is the BPR function, which models a link's cost as a non-linear function of the flow of vehicles on it (Bureau of Public Roads, 1964).

In the route choice process, drivers decide which route to take every day to reach their destinations. Usually, this process is modelled as a commuting scenario, where drivers' daily trips occur under approximately the same conditions. In this sense, each driver $i \in D$, with $|D| = d$, is modelled as an agent, which *repeatedly* deals with the problem of choosing the route that is expected to

[2] The app's travel information are also referred as *recommendations* along this paper.