Traffic Games with Incomplete Travel Information

Toshihiko Miyagi
Graduate School of Information Science, Tohoku University, Japan
toshi_miyagi@plan.civil.tohoku.ac.jp

Genaro C. Peque, Jr.
Graduate School of Information Science, Tohoku University, Japan
gpequejr@plan.civil.tohoku.ac.jp

Abstract

In this paper, we consider a traffic game where a group of self-interested agents tries to optimize their utility by choosing the route with the least travel time, and propose an ε-logit learning model that converges to an ϵ-Nash equilibrium (or a logit equilibrium) in the traffic game. The model consists of a N-person repeated game where the players know their strategic space and their realized payoffs, but are unaware of the information about the other players. The traffic game is essentially stochastic and described by stochastic approximation equations. An analysis of the convergence properties of the ε-logit learning rule is presented. Finally, with using a single origin-destination network connected by some overlapping paths, the validity of the proposed algorithms is tested.

Introduction

Traffic games having been considered in transportation research have been generally restricted to 2-person games: traveler versus nature, two travelers, traveler versus authority and so on.

The N-person game was usually formulated in somewhat unnatural way where a single origin-destination pair for trips takes role as a single player. The purpose of this paper is to study the route-choice behaviors of users in a traffic network comprised of a number of discrete, interactive decision-makers, which in turn implies that the traffic game considered here is a N-person non-cooperative game. In addition, our game-theoretical formulation of the traffic game is realistic and plausible in the sense that each agent doesn't know even his payoff (or cost) function as well as the other agents' information (payoffs, actions, strategies). The only information that each agent uses is the realized payoffs that are obtained by day-to-day travel experiences. This setting of the traffic game is referred to as a naive user problem. We can consider another type of traffic game where each agent knows his own payoff function.

This traffic game is referred to as an informed user problem because this behavioral assumption allows for each user to estimate the payoffs obtained if he had chosen other actions.

In the traditional user equilibrium, users are assumed to be infinitely divisible; users are assumed to be rational enough to know the existence of equilibrium; rationality is assumed to be common knowledge; all the data of the transportation systems are known to every user that is also common knowledge. We describe a user's route-choice behavior as repeated actions by an individual decision-maker with bounded rationality. Equilibrium is accounted for as a result of users' learning process from their day-to-day experiences. This kind of formulation naturally yields the question about what kind of equilibrium might arise as a consequence of a long-run non-equilibrium process of learning and adaptation.

The main objective of this paper is two folds: The first is to study the convergence properties of learning processes of the two types of traffic games. The second is to propose a unified learning algorithm that is applicable to both the naive user problem and the informed user problem within the same framework. Our approach is formally stated as stochastic traffic games where at each time period user's action (a pure strategy) is randomly determined according to his mixed strategy.

The Model

At the first, user $i \in \mathcal{I}$ is assumed to obtain noisy payoff $\tilde{u}_t^i$ at time t:

$$\tilde{u}_t^i := u^i(a_t^i, a_t^{-i}) + \varepsilon^i(a^i) \tag{1}$$

Where $\mathcal{I} = \{1, \ldots, i, \ldots N\}$: The set of users; a_t^i: The action selected by user i at time t, which is an element of the set of actions, A^i; $\varepsilon^i(a^i)$: The additive noise on observation of the realized payoff. The action specific average payoff at time t is expressed as:

$$\hat{u}_t^i(a^i) = \frac{1}{t z_t^i(a^i)} \sum_{s=0}^{t-1} \tilde{u}_s^i \mathbf{1}_{\{a_s^i = a^i\}} \tag{2}$$

where $z_t^i(a^i)$ denotes the empirical distribution visiting an action, defined as:

$$z_t^i(a^i) = \frac{1}{t} \sum_{s=0}^{t-1} \mathbf{1}_{\{a_s^i = a^i\}} \tag{3}$$

The indicator function $\mathbf{1}_{\{y\}}$ takes the value of 1 if the statement y is true, otherwise the value of zero. Since both $\hat{u}_t^i(a^i)$ and $z_t^i(a^i)$ are random variables, those are approximated by following equations:

$$\hat{u}_t^i(a^i) = \hat{u}_{t-1}^i(a^i) + \frac{1}{nz_t^i(a^i)}(\tilde{u}_{t-1}^i - \hat{u}_{t-1}^i(a^i))\mathbf{1}_{\{a_t^i = a^i\}} \tag{4}$$

$$z_t^i(a^i) = z_{t-1}^i(a^i) + \frac{1}{n}(\beta_{t-1}^i(a^i) - z_{t-1}^i(a^i)) \tag{5}$$

The mixed strategy (behavioral probability) of player i for an action is expressed in terms of the realized payoff that the user really experienced and the estimated payoffs of the other actions that he did not really choose;

$$\sigma^i(a^i) = \frac{\exp\left\{\hat{u}^i(a^i)/\mu^i\right\}}{\sum_{b^i \in A^i} \exp\left\{\hat{u}^i(b^i)/\mu^i\right\}} \tag{6}$$

We can also show that

$$\lim_{t\to\infty} \hat{u}_t^i(a^i) = u^i(a^i, z^{-i}) \tag{7}$$

This implies that the long run average of the realized payoff approaches to the payoff obtained when each user can observe the other users' behaviors. In other words, information a naïve user can use approaches to that of an informed user for a long time. This in turn implies that $\hat{u}_t^i(a^i)$ included in eqs.(4) and (6) can be replaced by $u^i(a^i, z_t^{-i})$ and that the behavioral probability become consistent with the best response. The asymptotic convergence property of this process is analyzed by the ODE approach. However, we adopt a slightly different model from the ordinary logit model that we refer to as a ε-logit model. The ε-logit is a linear combination of the ordinary logit and the probability one:

$$\beta_t^i(a^i) = \varepsilon\sigma_{t-1}^i(a^i) + (1-\varepsilon) \tag{8}$$

User i assigns the value given by (6) to an action included in the set of best responses defined as:

$$BR_t^i := \left\{a_*^i \in A^i : u_t^i(a_*^i) = \max_{a^i \in A^i} \hat{u}_t^i(a^i)\right\} \tag{9}$$

The rest of the $\beta_t^i(a^i)$ is assigned to the other actions.

It is ensured that our algorithm converges to a Nash distribution of the traffic game. If users know their payoff functions, then we can say more strong result. In this case, we can show that the learning algorithm converges to a pure Nash equilibrium (PNE) in congestion game in a relatively efficient way. One of the merits of the learning algorithm proposed here is that it can apply to both cases of the naive user and the informed user problem within the same framework.