Estimating learning dynamics in the routing game

Kiet Lam
UC Berkeley
kiet.lam@berkeley.edu

Walid Krichene
UC Berkeley
walid@eecs.berkeley.edu

Alexandre Bayen
UC Berkeley
bayen@berkeley.edu

ABSTRACT

The routing game models congestion on transportation and communication networks. We consider an online learning model of player dynamics: at each iteration, every player chooses an route (or a probability distribution over routes), then the joint decision of all players determines the costs of each path, which are then revealed to the players. We first review convergence guarantees of such online learning dynamics. Then, we consider the following estimation problem: given a sequence of player decisions and the corresponding costs, we would like to fit the learning model parameters to these observations. We consider in particular entropic mirror descent dynamics, and develop a numerical solution to the estimation problem.

We demonstrate this method using data collected from a routing game experiment: we develop a web interface to simulate the routing game. When players log in to the interface, they are assigned an origin and destination on the graph. They can choose, at each iteration, a distribution over their available routes, and each player seeks to minimize her own cost. We collect a data set using this interface, then we apply the proposed method to fit the learning model parameters. We observe in particular that after an exploration phase, the joint decision of the players remains within a small distance of the Nash equilibrium. We also use the estimated model parameters to predict the flow distribution over routes, and compare these predictions to the actual distribution. Finally, we discuss some of the qualitative implications of these findings, and directions for future research.

1. INTRODUCTION

The routing game is a non-cooperative game that models congestion in transportation and communication networks. The game is given by a directed graph that represents the network, and each player is given by a source node and destination node, and seeks to send traffic (either packets in a communication setting, or cars in a transportation setting) while minimizing the total delay of that traffic. The delay is determined by the joint decision of all players, such that whenever an edge has high load, it becomes congested and any traffic using that edge incurs additional delay.

This model of congestion is simple yet powerful, and routing games have been studied extensively since the seminal work of Beckman [1]. The Nash equilibria of the game are simple to characterize, and have been used to quantify the inefficiency of the network, using the price of anarchy [13]. However, the Nash equilibrium concept may not offer a good descriptive model of actual behavior of players, as argued by many authors, including for example [5]. Besides the assumption of rationality, which can be questioned, the Nash equilibrium assumes that players have a complete description of the structure of the game, their cost functions, and the cost functions of other players. This model is arguably not very realistic for the routing game, as one does not expect users of a network to have an accurate representation of the cost function on every edge of the network, or of the other users of the network. One alternative to the Nash equilibrium concept is a model of repeated play, sometimes called learning models or adjustment models. In such models, one assumes that each player makes decisions iteratively, and uses the outcome of each iteration to adjust their next decision. Formally, if $x_k^{(t)}$ is the decision of player k at iteration t, and $\ell_k^{(t)}$ is the corresponding vector of costs, then player k faces a sequential decision problem in which she chooses $x_k^{(t)}$ then observes $\ell_k^{(t)}$. These sequential decision problems are coupled through the cost functions, since $\ell_k^{(t)}$ depends not only on $x_k^{(t)}$ but also on $x_{k'}^{(t)}$ for $k' \neq k$ (but players do not neces-

sarily model this coupling). Such models have a long history in game theory, and date back to the work of Hannan [7] and Blackwell [2].

When designing a model of player decisions, many properties are desirable. Perhaps the most important property is that the dynamics should be consistent with the equilibrium of the game, in the following sense: asymptotically, one should expect the learning dynamics to converge to the equilibrium of the full information, one-shot game (be it Nash equilibrium or other, more general equilibrium concepts). In this sense, players "learn" the equilibrium asymptotically. Much progress has been made in recent years in characterizing classes of learning dynamics which are guaranteed to converge to an equilibrium set [6, 9, 8, 5]. In particular for the routing game, different models of learning have been studied for example by [4, 3, 10, 12, 11].

In this paper, we briefly review some of the known models of learning in the routing game. We focus in particular on the mirror descent model used in [11]. This model describes the dynamics of the learning as solving, at each step, a simple minimization problem parameterized by a learning rate. Formally, the decision at iteration $t+1$ is obtained by solving

$$x^{(t+1)}(\eta) = \arg\min_{x_k \in \Delta^{\mathcal{A}_k}} \left\langle x, \ell_k^{(t)} \right\rangle + \frac{1}{\eta} D_\psi(x_k, x_k^{(t)}),$$

where ψ is a distance generating function with corresponding Bregman divergence D_ψ, and $\eta^{(t)}$ is a learning rate. The model is reviewed in detail in Section 2. Intuitively, the learning rate $\eta_k^{(t)}$ determines how aggressive the player is when updating her strategy: a small learning rate results in a small change in strategy (i.e. $x_k^{(t+1)}$ is close to $x_k^{(t)}$), while a large learning rate results in a significant change.

Motivated by this interpretation of learning rates, we propose in Section 3, the following estimation problem: given a sequence of player decisions $(x_k^{(t)})_t$, and the sequence of corresponding losses $(\ell_k^{(t)})$, can we fit a learning model to these observations? A simple approach is to assume that the player is using a given distance generating function ψ, and estimate η for example by minimizing the distance between the observed decision $\bar{x}^{(t+1)}$, and the decision predicted by the model, $x^{(t+1)}(\eta)$. More precisely, we can choose $\eta^{(t)}$ to minimize $D_\psi(\bar{x}^{(t+1)}, x^{(t+1)}(\eta))$. We show that in the entropic case (when ψ is the negative entropy), this problem is convex, thus $\eta^{(t)}$ can be estimated efficiently e.g. by using gradient descent. This method allows us to estimate one parameter $\eta^{(t)}$ per iteration t. When we have a sequence of observations available, it can be desirable to control the complexity of the model by assuming a parameterized sequence of learning rates. Thus, we propose a second method which assumes that the learning rate is of the form $\eta^{(t)} = \eta^{(0)} t^{-\alpha}$, with $\alpha \in (0, 1)$. The resulting estimation problem is non-convex in general, but since it is a two dimensional problem, it can be solved efficiently. Finally, we consider a family of distance generating functions ψ_ϵ, parameterized by ϵ, that can be viewed as a generalization of the negative entropy function. These generalized entropy functions also offer some desirable properties that we discuss in Section 3. We also briefly discuss potential uses for the estimated model: for example, the model can be used simply to predict the decision of the players over the next few iterations, by propagating the model forward with the estimated values of the parameters; more generally, the model can be used to formulate a receding-horizon optimal control problem, by using the current estimate of the model as a plant in the control problem.

In the second part of the paper, we present an experimental setting which we developed to collect data on routing decisions. We developed a web interface in which a master user can create an instance of the routing game by defining a graph and cost functions on edges of the graph. Then other users can connect to the interface as players. The game then proceeds similarly to our learning model: at each iteration, every player chooses a flow distribution on their available routes (using a graphical interface with sliders), then their decisions are sent to a backend server, which computes the total cost of each route, and sends back this information to each player. In Section 4, we describe the experimental setting, some implementation details, as well as the nature of the collected data. We then use this data to run the estimation tasks which were proposed in Section 3, and give some qualitative and quantitative insights into the behavior of players. In particular, we observed that in the first few iterations, the flow distributions oscillate, which corresponds to a high value of estimated learning rates. For later oscillations, the flow distributions are, in general, close to equilibrium, and the learning rates are lower, although some players may occasionally move the system away from equilibrium by performing an aggressive update (corresponding to a high learning rate). We also solve the decay rate estimation problem

2. THE ROUTING GAME AND THE LEARNING MODEL

2.1 The routing game

2.2 The learning model

2.3 Mirror descent dynamics and brief review of convergence results

3. LEARNING MODEL ESTIMATION

3.1 Estimating a single learning rate

3.2 Estimating the decay rate of the learning rate sequence

3.3 A parameterized family of distance generating functions

3.4 Application to prediction and receding horizon control

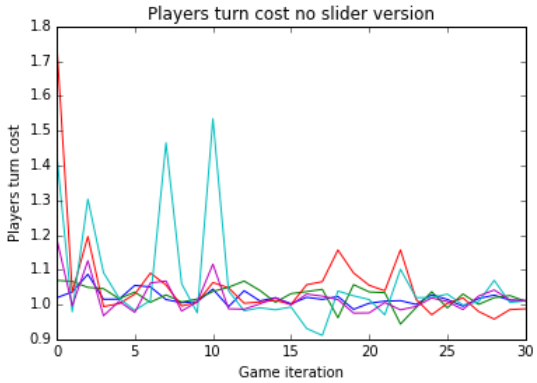

Figure 1: Player normalized cost for no slider version.

4. EXPERIMENT

4.1 I

Two versions of the game were presented to the user. One version of the game allowed each player to select the flow distribution for each path, the other version of the game only allowed the player to choose an "exploration" parameter, which selects the flow distribution for the players based on the KL-divergence. The players interfaced with the game through a web interface where they can enter their inputs and receive feedback of each of their path cost and the cumulative cost.

The cost for each player's distribution were normalized by the calculated equilibrium flow to account for imbalances in the network.

4.2 II

In the first version of the game, each player were allowed sliders to change how much of the flow are distributed to each path. All paths for a player have the same origin and destination node pair, but each player had unique origin and destination node pair.

We estimate the learning rate, $\hat{\eta}_t^k$, of the players at turn t by

$$\hat{\eta}_t^k = \arg\min_{\eta \geq 0} D_{\psi_k}(\hat{x}_k^{(t+1)}(\eta), x_k^{(t+1)})$$

From these estimated learning rate sequence, we then estimate the flow distribution of iteration $t+1$ using $\hat{\eta}_t^k$ and $x_k^{(t)}$ with 1.

4.3 III

In the second version of the game, each player were only to change a single parameter η. This parameter controls how much each player want to explore from their previous flow distribution according to

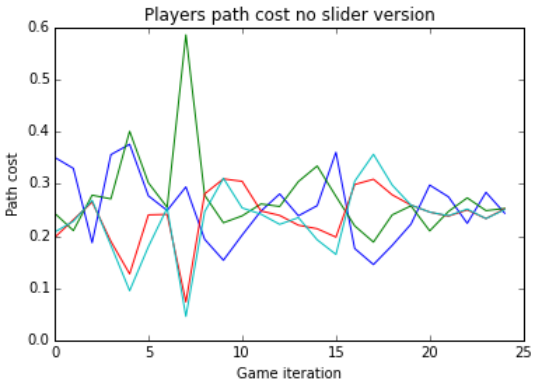

Figure 2: Predicted path cost.

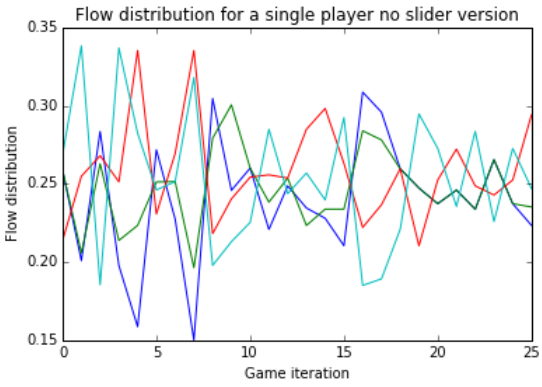

Figure 3: Player flow distribution.

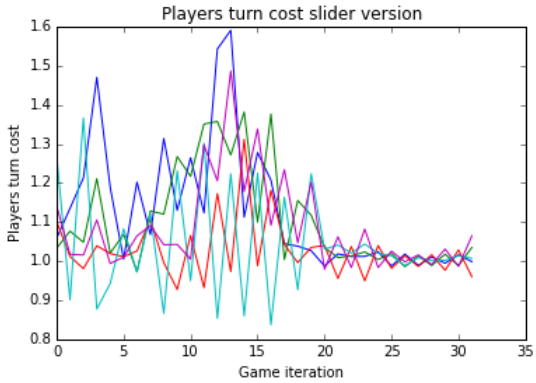

Figure 4: Player normalized cost for slider version.

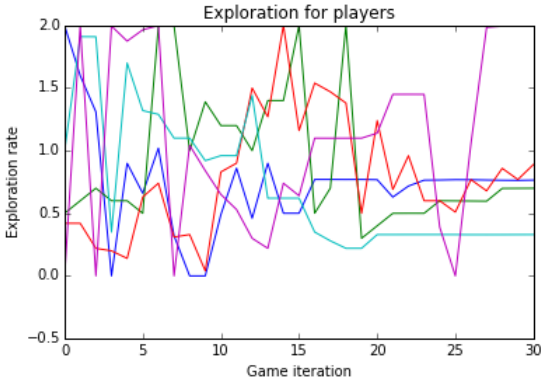

Figure 5: Player exploration rate.

$$(x_k^{(t+1)})_i = \frac{e^{-\eta_t^k(\ell_k^{(t)})_i}}{\sum_j (x_k^{(t)})_i e^{-\eta_t^k(\ell_k^{(t)})_i}} \quad (1)$$

. The parameter for each turn is shown in 5. We use a small ϵ to ensure that we will have no underflow for $x_k^{(t)}$ for numerical purposes. Then **??** becomes

$$(x_k^{(t+1)})_i = \frac{e^{-\eta_t^k(\ell_k^{(t)})_i}}{\sum_j (x_k^{(t)})_i e^{-\eta_t^k(\ell_k^{(t)})_i}} \quad (2)$$

5. CONCLUSIONS

Conclusion goes here.

Acknowledgments

Acknowledgement goes here.

6. REFERENCES

[1] M. J. Beckmann, C. B. McGuire, and C. B.

Winsten. Studies in the economics of transportation. 1955.

[2] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[3] A. Blum, E. Even-Dar, and K. Ligett. Routing without regret: on convergence to nash equilibria of regret-minimizing algorithms in routing games. In *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, PODC '06, pages 45–52, New York, NY, USA, 2006. ACM.

[4] S. Fischer and B. Vöcking. On the evolution of selfish routing. In *Algorithms–ESA 2004*, pages 323–334. Springer, 2004.

[5] M. J. Fox and J. S. Shamma. Population games, stable games, and passivity. *Games*, 4(4):561–583, 2013.

[6] Y. Freund and R. E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1):79–103, 1999.

[7] J. Hannan. Approximation to Bayes risk in repeated plays. *Contributions to the Theory of Games*, 3:97–139, 1957.

[8] S. Hart. Adaptive heuristics. *Econometrica*, 73(5):1401–1430, 2005.

[9] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98(1):26 – 54, 2001.

[10] R. Kleinberg, G. Piliouras, and E. Tardos. Multiplicative updates outperform generic no-regret learning in congestion games. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 533–542. ACM, 2009.

[11] S. Krichene, W. Krichene, R. Dong, and A. Bayen. Convergence of heterogeneous distributed learning in stochastic routing games. In *53rd Allerton Conference on Communication, Control and Computing*, 2015.

[12] W. Krichene, B. Drighès, and A. Bayen. Learning nash equilibria in congestion games. *SIAM Journal on Control and Optimization (SICON)*, 2015.

[13] T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2):236–259, 2002.

APPENDIX

Appendix goes here.