Learning Across Repeated Congestion Games

SAM GRIESEMER* and CONNOR MONAHAN*, Washington University in St. Louis

Congestion games are a class of games where agents must select from a set of resources whose costs are determined by their demand. Traffic networks are a common example from this class of games, where resources are roads whose cost is dependent on the number of agents driving along that road. In this project, we explore the actions of agents over time in a simple road network environment. We implement three different learning algorithms that allow the agents to make inferences about their environment and act rationally according to their beliefs learned across repetitions of the traffic game. We analyze the dynamics of each of these learning strategies and their impact on the behavior of the agents. We additionally report on the agents' convergence to equilibrium, and how Braess' paradox impacts agent behavior in the traffic network.

1 INTRODUCTION

Congestion games are games with resources that take on a value inversely proportional to the number of agents that use them. This class of games naturally lends itself to problems involving network flow, where the value of following a particular path through a network decreases as flow increases along that path. One important example of this type of game is traffic networks, which model traffic flow through a network of roads. If agents in such a system can be realistically modeled, then we can better understand the dynamics of real traffic networks and predict the flow of traffic.

There are a number of interesting phenomena that arise when considering the structure of road networks. One such phenomenon is Braess' paradox, which is commonly defined as the observation of a negative impact to overall traffic flow when roads are added to a road network. This paradox arises from the typically well-intentioned act of adding roads (e.g. highways) to increase the flow of traffic throughout a road network. Intuitively we might expect this outcome; more roads provide more travel options for drivers and seemingly better distribute traffic, thus reducing total travel time for everyone. However, there are certain instances where the opposite occurs, and adding a road instead increases overall traffic time for all drivers. This is often a result of the new road presenting a better alternative over existing routes, incentivizing all drivers to take the new road. The newly added route then becomes expensive for all drivers, none of which now have incentive to switch back to pre-existing routes. We see a canonical example of Braess' paradox in the setup for our problem, described in Section 1.1.

In this project, we explore the behavior of agents in repeated congestion games by simulating a number of agent learning algorithms on the simple traffic network described in Section 1.1. We test three different learning strategies: fictitious play, ϵ-greedy exploration, and Thompson sampling. These algorithms define a learning strategy and action mechanism that allows each agent to maintain beliefs about the environment as repeated iterations of the traffic congestion game are played out. These strategies are each explained in further detail in Section 2.

1.1 Setup

When developing a framework for simulating agents on a road network, we used the general road structure and costs as depicted in Figure 1.

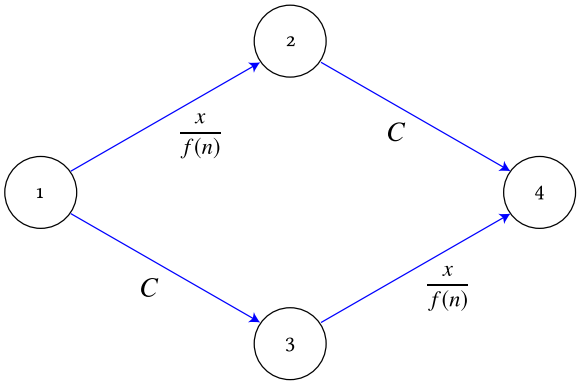

Fig. 1. Network diagram without the superhighway

Here the values along each of the blue paths represent the costs incurred by the agents traversing those roads. The value C is defined as some fixed cost not dependent on the number of agents traversing the road. For the roads with a cost given by $\frac{x}{f(n)}$, agents traveling on these roads each incur a cost proportional to x, the total number of agents traveling along that road. The constant of proportionality is defined by some function $f(n)$, where n is the total number of agents in the simulation. We present the costs in this way to ensure roads with dynamic costs are properly weighted across simulations with varying numbers of agents.

In Figure 2, we introduce an additional road referred to as the "superhighway". This road connects nodes 2 and 3 at a fixed cost of zero for all agents that travel on it. This is the canonical Braess'

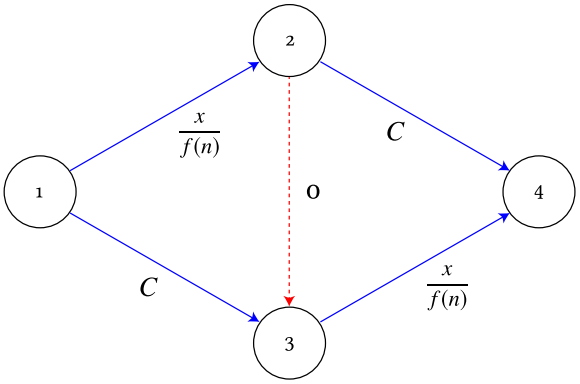

Fig. 2. Network diagram including superhighway as indicated by the dashed red line

*Both authors contributed equally to this research. This project follows option 2: repeated congestion games.

Authors' address: Sam Griesemer, samgriesemer@gmail.com; Connor Monahan, cmonahan@wustl.edu, Washington University in St. Louis, 1 Brookings Dr. St. Louis, Missouri, 63367.

paradox network structure mentioned previously, which as we will later observe shifts the network equilibrium (for the worse). Note that in order for Braess' paradox to have an effect after adding a superhighway, these costs must be tuned to the number of agents. We use $f(n) = n/20$, evaluation to 100 for 2,000 agents (used commonly across experiments in Section 3).

2 METHODS

When simulating agents in the traffic environment, we experimented with fictitious play, ϵ-greedy exploration, and Thompson sampling for our agent learning strategies. In this section we briefly describe each of these methods and provide relevant implementation details.

2.1 Fictitious Play

Fictitious play is a learning strategy where agents assume their opponents choose actions according to a stationary mixed strategy at each iteration of the game [1]. An agent computes an opponent's assumed mixed strategy using their empirical distribution of actions marginalized over time. That is, the marginal empirical distribution of agent i at time t is given by

$$p_i^t(a_i) = \frac{1}{t}\sum_{\tau=0}^{t-1} [\![a_i^\tau = a_i]\!],$$

where $a_i \in A_i$ is some action in agent i's action space A_i. This is the mixed strategy each agent other than i will assume agent i is playing at time t. Each agent i then computes the product of their opponents' empirical distributions $\mathbf{p}_{-i}^t$ (which is defined over $\prod_{j\neq i} A_j$). The distribution $\mathbf{p}_{-i}^t$ serves as a proxy for the true joint distribution over agent i's opponents' strategy profiles. Each agent i then selects the action that best responds to $\mathbf{p}_{-i}^t$, i.e. the action for which they have the minimum expected cost under the joint distribution of their opponents assumed strategies.

We note that fictitious play is classically defined for 2-player games, and converges on 2-player potential games as shown by [2]. For this reason, along with the computational difficulty of scaling the policy to a large number of agents, we restrict our fictitious play experiments (Section 3.1 to a relatively low number of agents ($n = 2, \ldots, 6$).

2.2 ϵ-greedy

ϵ-greedy is a multi-armed bandit strategy that repeats the best action seen thus far $(1-\epsilon)$% of the time and deviates to a randomly-selected action ϵ% of the time [4]. In our experiments (Section 3.2), we decided to use an epsilon value of 0.01 for 2,000 agents. Thus the expected number of agents deviating to a randomly selected action is 20 for each iteration. The entire history for each agent was stored for reference when selecting the best action. In the first iteration of the game, agents pick their actions uniformly at random from the action space.

2.3 Thompson Sampling

Thompson sampling takes a Bayesian approach to learning agent behavior, using a prior, likelihood, and posterior which are updated using Bayes' rule from past experiences. Every time the agent is required to take an action, it samples a value θ^* from the posterior and uses that to choose the action maximizing the expected likelihood given those parameters [3].

Our reward signal was constructed from the agent's individual cost of traversing their last road. This cost is distributed according to a multinomial distribution. Each agent has a choice between 2 or 3 roads (depending on the presence of a superhighway), which determines the number of categories of our distribution. Our prior and posterior were chosen to be Dirichlet-distributed due to its conjugacy with respect to a multinomial likelihood and extensive flexibility. The parameters of the Dirichlet prior were set initially so as to create a uniform prior over every action/route. The posterior was updated at each iteration using Bayes' rule by adding the reward signals to the appropriate action position in the α vector.

3 EXPERIMENTS AND RESULTS

Here we report and describe results for a number of different questions for each of the learning strategies. Questions we set out to answer by our observations include (1) how outcomes vary on the traffic network with and without the superhighway, (2) how the number of agents affects the simulation outcome (along with factors like rate of convergence), and (3) how different initial settings and/or prior beliefs affect outcomes. Finally, we compare the results across all of the learning strategies and draw conclusions about performance on this traffic environment.

3.1 Fictitious Play

For the reasons given in Section 2.1, we only ran experiments with 2, 4, and 6 agents. For each of these cases, we noticed consistent convergence to the expected equilibrium for both traffic networks with and without the superhighway.

That is, without the superhighway, all experiments eventually reached the equilibrium outcome where each route is taken by half of the agents. We note that, if started in equilibrium, no agent would ever switch their action to taking a different path at any future iteration. When experiments were started with an uneven number of agents taking either route, we observed that agents would initially synchronize their actions, selecting the same route at each repetition. This is because of the two available routes appeared to be the best choice for each agent, and thus was taken by all agents. Over time, however, the simple stochasticity enforced by fictitious play on indifference steps in and randomly varies agents' actions. This ultimately results in acheived equilibrium.

With the superhighway enabled, all experiments had an identical process. Regardless of the agents' starting routes, after one iteration of the game all agents would be taking the superhighway. This is due to the stationary assumptions each agent is making about their opponents, and the easily identifiable benefit each agent has by greedily switching to the superhighway under these assumptions.

All in all, our observations matched our expectation convergence to equilibrium in accordance with the proof of fictitious play's convergence for two-player potential games in [2].

3.2 ϵ-greedy

Having $n = 2000$ agents make decisions using a ϵ-greedy policy, we observe the total cost and road choice for all agents as shown in

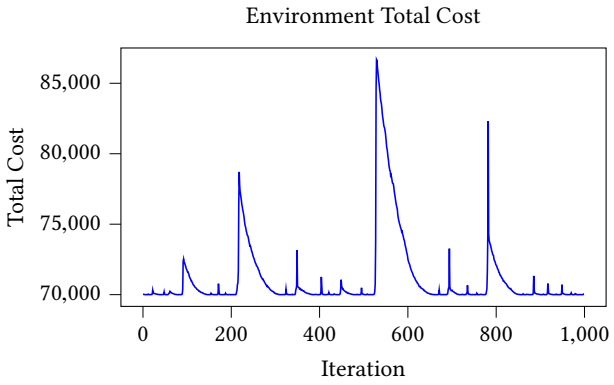

Fig. 3. 2000 Agents using ϵ-greedy without superhighway

Fig. 4. 2000 Agents using ϵ-greedy without superhighway

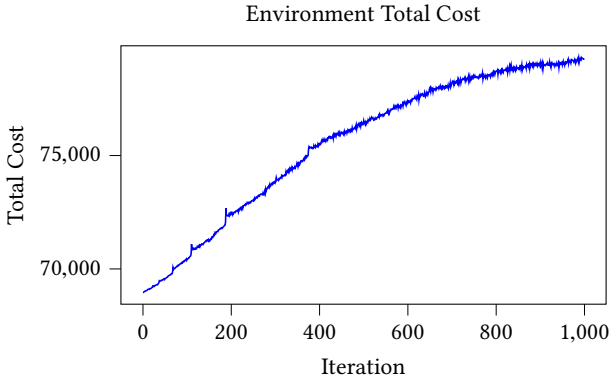

Fig. 5. 2000 Agents using ϵ-greedy with superhighway

Figure 3 (before adding the superhighway). The corresponding plot showing the number of agents taking each action at each iteration can be seen in 4. After adding the superhighway, the results change to those shown in Figure 5. Again, the corresponding visualization of the number of agents taking each of the possible actions is shown in Figure 6. Both of these experiments appear to converge after many iterations, although there is a noticeable pattern of large divergences in the case where a superhighway is not available. We can see Braess' paradox in action: after adding the superhighway, the total cost experienced by every agent increased at convergence.

These results, especially in comparison with those from other methods, beg the question: what improvement can changing ϵ have

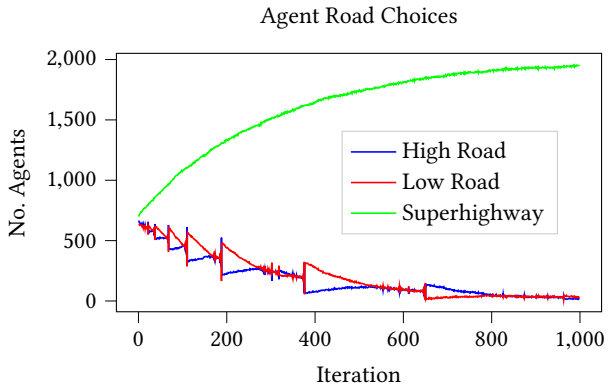

Fig. 6. 2000 Agents using ϵ-greedy with superhighway

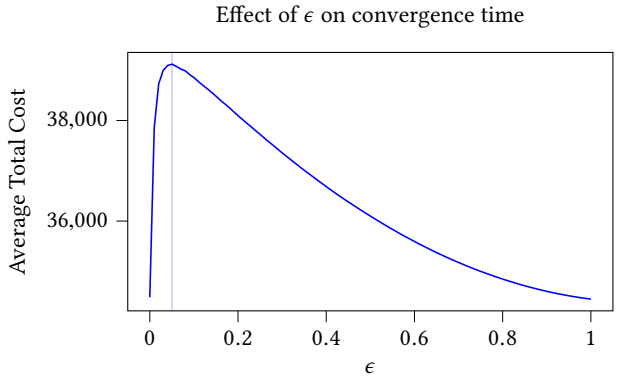

Fig. 7. Evaluating convergence time for values of ϵ

on convergence? We then varied epsilon for 100 unique values between 0 and 1 with the superhighway and plotted the average total score (a proxy for area under total cost curve) in Figure 7. These results varying prior parameters suggest that $\epsilon = 0.05$ lead to the quickest convergence when using the superhighway as this maximizes the curve. New graphs using this value are available in appendix A and demonstrate faster convergence than previously shown with the ϵ-greedy strategy. We additionally repeated this experiment using an environment with a lower number of agents ($n = 100$) and noticed that the time to convergence decreased, as we might expect.

3.3 Thompson Sampling

Having all agents make decisions using a Thompson sampling policy, the behavior before adding a superhighway can be seen in Figure 8. This plot demonstrates the relatively quick convergence to equilibrium, at a total cost of 70,000. Figure 9 shows the number of actions that take each route at each iteration.

After adding the superhighway, the results change to those presented in Figure 10. We can see here a very rapid convergence: after less than 200 iterations, the total cost stabilizes and individual agents stick to their proven exploitative behavior, as shown in Figure 9. Agents appear to distribute themselves between the two roads in the trial without a superhighway, as is optimal. After adding a superhighway, they quickly all rush to use the superhighway as shown in Figure 11, since it optimizes their individual costs. We additionally

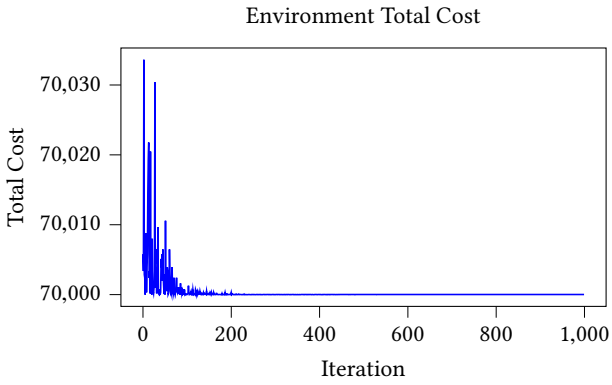

Fig. 8. 2000 Agents using Thompson sampling without superhighway

Fig. 9. 2000 Agents using Thompson sampling without superhighway

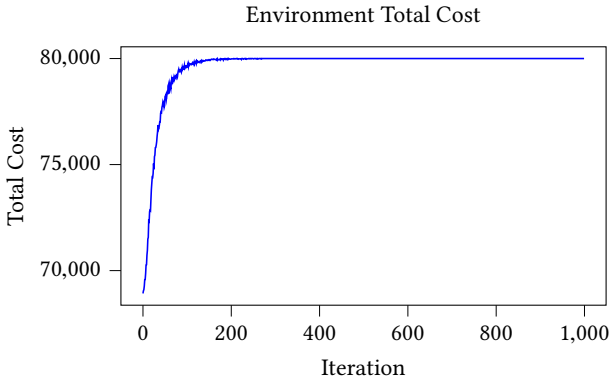

Fig. 10. 2000 Agents using Thompson sampling with superhighway

repeated this experiment on larger and smaller bodies of agents, and observed that the time to convergence remained approximately the same. As a result, we conclude that the rate of convergence when Thompson sampling is used in this environment does not appear to depend heavily on the number of agents involved in the simulation.

4 CONCLUSION

Among all of the learning methods we experimented, Thompson sampling appeared to move the quickest toward convergence of the agent-optimal cost in the traffic network environment. It also exploited the optimal solution the most after finding it, not being prone to the large cost spikes caused by poor exploration as

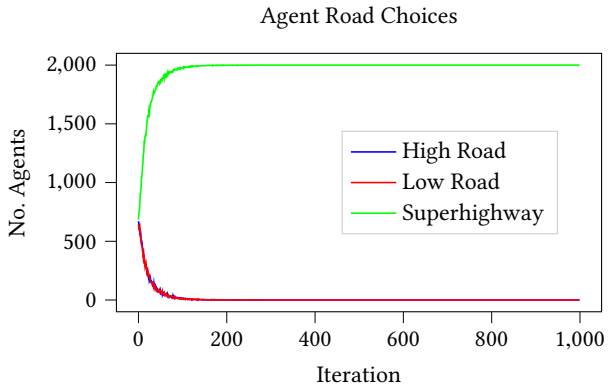

Fig. 11. 2000 Agents using Thompson sampling with superhighway

seen in ϵ-greedy (Figure 3). It additionally did not require extensive prior tuning to perform effectively, only working from an uninformative, uniform prior. Both Thompson sampling and ϵ-greedy approaches are more applicable to games with large numbers of agents when compared to fictitious play with our implementation. All tested strategies were shown to make individually rational decisions at convergence, as shown by every agent eventually taking the superhighway in that case. All strategies also settled towards a half-and-half split between the roads in the case that there was no superhighway available.

5 FUTURE WORK

The success of Thompson sampling in this environment suggests that alternative Bayesian methods of creating a prior over actions and updating it in response to observations may also be applicable to this problem. Bayesian optimization allows use of a wide variety of policy options such as expected improvement to select the point (or action in this case) to explore next. However, each action selected here would include an indeterminate amount of noise reflecting the actions of other individuals.

REFERENCES

[1] Ramesh Johari. 2007. Fictitious play. Lecture notes. http://web.stanford.edu/~rjohari/teaching/notes/336_lecture6_2007.pdf
[2] Dov Monderer and Lloyd S Shapley. 1996. Potential games. *Games and economic behavior* 14, 1 (1996), 124–143.
[3] Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen, et al. 2018. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning* 11, 1 (2018), 1–96.
[4] R.S. Sutton and A.G. Barto. 1998. *Reinforcement learning: An introduction*. Vol. 116. Cambridge Univ Press.

A GRAPHS AFTER PRIOR OPTIMIZATION

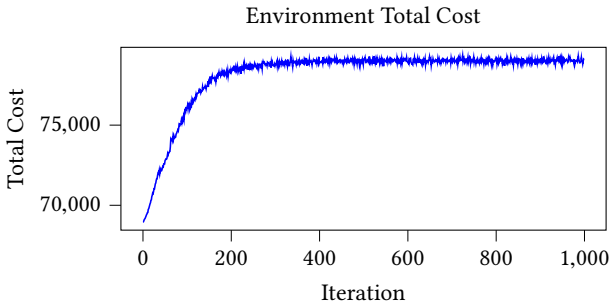

Fig. 12. 2000 Agents using ϵ-greedy with superhighway

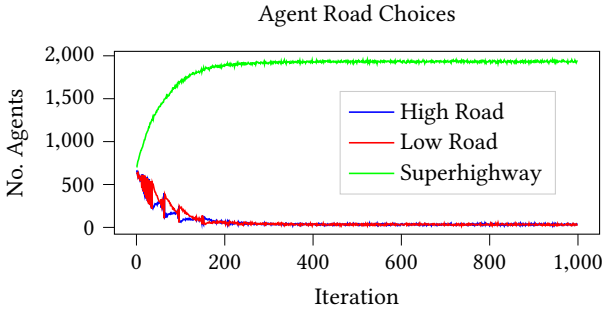

Fig. 13. 2000 Agents using ϵ-greedy with superhighway