Modeling Bounded Rationality of Agents During Interactions

(Extended Abstract)

Qing Guo and Piotr Gmytrasiewicz
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607
{qguo, piotr}@cs.uic.edu

ABSTRACT

In this paper, we propose that bounded rationality of another agent be modeled as errors the agent is making while deciding on its action. We are motivated by the work on quantal response equilibria in behavioral game theory which uses Nash equilibria as the solution concept. In contrast, we use decision-theoretic maximization of expected utility. Quantal response assumes that a decision maker is approximately rational, i.e., is maximizing its expected utility but with an error rate characterized by a single error parameter. Another agent's error rate may be unknown and needs to be estimated during an interaction. We show that this error rate can be estimated using Bayesian update of a suitable conjugate prior, and that it has a sufficient statistic of fixed dimension under strong simplifying assumptions. However, if the simplifying assumptions are relaxed, the quantal response does not admit a finite dimensional sufficient statistic, and a more complex update is needed.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents, Multiagent systems*; I.2.6 [**Artificial Intelligence**]: Learning—*Parameter learning*

General Terms

Theory, Design

Keywords

Bounded rationality, Multi-agent interaction, Multi-agent learning, Formal models of agency

1. INTRODUCTION

In AI, an agent's (perfect) rationality is defined as the agent's ability to execute actions that, at every instant, maximize the agent's expected utility, given the information it has acquired from the environment [8]. Modeling others as rational has a long tradition in AI and game theory, but modeling other agents' departures from rationality is difficult and controversial. This paper builds on an approach to modeling bounded rationality called quantal response [2, 6, 7]. It is a simple model which uses a single error parameter. Quantal response does not attempt to model the procedures, and their possible limitations, the agent may use to decide on its action; instead, it abstracts away the unobservable parameters specific to implementation and treats them as *noise* which produces non-systematic departures from perfect rationality. Quantal response specifies the probabilities of an agent's actions with the logit function of their expected utilities and the agent's error parameter, λ. Thus actions that are suboptimal are possible, but their probabilities increase with their expected utilities. Usually, an agent's error parameter is not directly observable and must be inferred by observing its behavior. We take a Bayesian approach to this and propose that the modeling agent maintain a probability distribution over possible values of λ for the modeled agent, and that this probability be updated when new actions are observed. This paper shows that, in simple special cases, the error rate admits a sufficient statistic of fixed dimension, and thus there exist conjugate prior families for these cases; however, in more general cases, there is no finite dimensional sufficient statistic and no conjugate prior over λ.

2. LOGIT QUANTAL RESPONSE

For simplicity, we assume that a modeling agent, called i, is considering the behavior of one other agent, j. The logit quantal response is defined as follows [2, 6, 7]:

$$P(a_j) = \frac{e^{\lambda u_{a_j}}}{\sum_{l=1}^{m} e^{\lambda u_{a_l}}}, \tag{1}$$

where $\{a_l : l = 1, 2, ..., m\}$ is a set of possible actions of agent j. $P(a_j)$ is the probability of agent j taking action a_j. $u_{a_j} \in \mathbb{R}$ is the expected utility of action a_j to agent j and $\lambda \geq 0$ is the (inverse) error rate of agent j. λ represents how rational agent j is: greater λ makes it more likely that j takes actions with higher utilities. When $\lambda \to +\infty$, $P(a_j) = 1$ for the action with the highest expected utility[1] and $P(a_j) = 0$ for all other actions. When $\lambda = 0$, $P(a_j) = 1/m$, $\forall j = 1, 2, ..., m$, which means agent j chooses actions at random.

Usually the error rate λ of agent j is not directly observable to agent i. Bayesian approach allows agent i to learn

[1]If there are many, say h, optimal actions with the same expected utilities, then $P(a_j) = 1/h$ for each of them.

Cite as: Modeling Bounded Rationality of Agents During Interactions (Extended Abstract), Qing Guo and Piotr Gmytrasiewicz, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 1285-1286.

this rate during interactions. To do this agent i needs a prior distribution, $f(\lambda)$, which represents i's current knowledge about agent j's error rate, and to observe agent j's action, a_j at the current step. The updated distribution is:

$$f(\lambda|a_j) = \frac{P(a_j|\lambda)f(\lambda)}{\int_0^\infty P(a_j|\lambda')f(\lambda')\,d\lambda'}. \tag{2}$$

Equation (2) may not be easy to apply because repeatedly updating $f(\lambda)$ makes its form more and more complicated. To overcome this it is convenient to look for a conjugate prior family. In Bayesian probability, if the posterior distribution is in the same family as the prior distribution, then this prior is called a *conjugate prior* [3, 4]. Conjugate priors are convenient because one just needs to update the parameters of the conjugate prior distribution (hyperparameters) to realize the Bayesian update.

3. STATIC EPISODIC ENVIRONMENTS

We first consider the simplest case, when agent j's expected utilities u_{a_l} for all actions are *known* to agent i and remain the same during the interaction. In other words, agent j is not updating his beliefs since the environment is static and episodic [8] and i is observing j acting in the same decision-making situation repeatedly. The derivation below follows techniques in [3, 4].

Consider the following family of distributions over λ:

$$f(\lambda; u, n) = \frac{e^{\lambda u}/(\sum_{l=1}^m e^{\lambda u_{a_l}})^n}{\int_0^\infty e^{\lambda' u}/(\sum_{l=1}^m e^{\lambda' u_{a_l}})^n\,d\lambda'}, \tag{3}$$

where n and u are hyperparameters. Here $n = 0, 1, ...,$ and u is restricted by $u < n \max_l u_{a_l}$. (3) is a valid probability density function since integral of the denominator converges if and only if $u < n \max_l u_{a_l}$. We claim that the family of distributions $f(\lambda; u, n)$ in (3) is a conjugate family of distributions over λ in static episodic environments with known utilities of actions. It can be proven that the update of the hyperparameters of this conjugate prior after observing that agent j executed his action a_j, with expected utility u_{a_j} is:

$$f(\lambda; u, n) \xrightarrow{a_j} f(\lambda; u + u_{a_j}, n + 1). \tag{4}$$

One can verify that once there is a valid prior, all the posteriors are always valid.

4. SEQUENTIAL DYNAMIC ENVIRONMENTS

We extend our approach to more complex case of dynamic sequential environment [8]. Again, we assume that expected utilities of j's actions are known to i, but now, since agent j may be updating his beliefs, the expected utilities of his actions do not remain constant but can take a finite number of values. We refer to each of the beliefs of agent j, together with his payoff function and other elements of his POMDP, as j's type, θ_j. Thus, the set of possible types of agent j, Θ_j, has K possible elements $1, 2, ..., K$. We denote $U(a_j|\theta_j = k) = u_{a_j,k}$, where $k = 1, 2, ..., K$, and assume that index k is observable (or computable) by agent i. Then the logit quantal response (1) for the probability of agent j taking action a_j given his kth type is:

$$P(a_j|k, \lambda) = \frac{e^{\lambda u_{a_j,k}}}{\sum_{l=1}^m e^{\lambda u_{a_l,k}}}. \tag{5}$$

Now Consider the following family of distributions:

$$f(\lambda; u, n_1, n_2, ..., n_K) = \frac{e^{\lambda u}/\prod_{k=1}^K(\sum_{l=1}^m e^{\lambda u_{a_l,k}})^{n_k}}{\int_0^\infty e^{\lambda' u}/\prod_{k=1}^K(\sum_{l=1}^m e^{\lambda' u_{a_l,k}})^{n_k}\,d\lambda'}, \tag{6}$$

where $n_k = 0, 1, ..., \forall k = 1, ..., K$; $u < \sum_{k=1}^K (n_k \max_l u_{a_l,k})$. (6) is valid since integral of the denominator converges if and only if $u < \sum_{k=1}^K (n_k \max_l u_{a_l,k})$. We claim that the family of distributions $f(\lambda; u, n_1, n_2, ..., n_K)$ in (6) is a conjugate family of distributions over λ in a sequential dynamic environment with perfect observability of finite number of types. The update of the hyperparameters of this conjugate prior:

$$f(\lambda; u, n_1, n_2, ..., n_K) \xrightarrow{a_j,k} f(\lambda; u + u_{a_j,k}, n_1, n_2, ..., n_{k-1}, n_k + 1, n_{k+1}, ..., n_K). \tag{7}$$

Once there is a valid prior, all the posteriors are always valid.

Now let us consider an even more general case, in which the expected utilities u_{a_l} are not limited to a finite number of values but can lie in some interval or even on the real line:

$$P(a_j|\boldsymbol{u}, \lambda) = \frac{e^{\lambda u_{a_j}}}{\sum_{l=1}^m e^{\lambda u_{a_l}}}, \tag{8}$$

where $u_l < u_{a_l} < u_l{}', l = 1, 2, ..., m$, $u_l \geq -\infty$ and $u_l{}' \leq \infty$ are lower and upper bounds of the expected utilities u_{a_l}, and where $\boldsymbol{u}$ is a vector of expected utilities of all m actions, $\boldsymbol{u} = (u_{a_1}, u_{a_2}, ..., u_{a_m})$. Again assume u_{a_l} are known to agent i, and he observes agent j's action a_j.

Forming a conjugate prior in this case may be impossible. The reason is that the construction of conjugate prior distributions [3, 4] is based on the existence of sufficient statistics of fixed dimension for the given likelihood function (equation (8)). However, under very weak conditions, the existence of fixed dimensional sufficient statistic restricts the likelihood function to the exponential family [1, 5]. Unfortunately, (8) does not belong to the exponential family when $m \geq 2$.

5. REFERENCES

[1] O. Barndorff-Nielsen and K. Pedersen. Sufficient data reduction and exponential families. *Math. Scand.*, 22:197–202, 1968.

[2] C. F. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, 2003.

[3] M. H. DeGroot. *Optimal Statistical Decisions (Wiley Classics Library).* Wiley-Interscience, April 2004.

[4] D. Fink. A Compendium of Conjugate Priors. Technical report, 1997.

[5] D. A. S. Fraser. On sufficiency and the exponential family. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(1):115–123, 1963.

[6] R. McKelvey and T. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10:6–38, 1995.

[7] R. McKelvey and T. Palfrey. Quantal response equilibria for extensive form games. *Experimental Economics*, 1:9–41, 1998.

[8] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach (Third Edition).* Prentice Hall, 2010.