Learning: Bayesian Learning

(Based on Shoham and Leyton-Brown (2008). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, Cambridge.)

Leen-Kiat Soh

Rational Learning is also sometimes called Bayesian Learning

What is rational? What is utility-maximizing?
(Think about “a better option” vs. “the best option”)

Introduction

- *Rational learning* adopts the same general model-based scheme as fictitious play
  - However, it allows players to have ***a much richer set of beliefs*** about opponents' strategies.
  - As in fictitious play, each player begins the game with some prior beliefs
  - After each round, the player uses ***Bayesian updating*** to update these beliefs.

Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Compute posterior probabilities by using likelihood and collecting data on priors

Likelihood
How probable is the evidence given that our hypothesis is true? (Computable from historical data)

Prior
How probable **was** our hypothesis before observing the specific evidence? (Computable from historical data)

$$P(H|e) = \frac{P(e|H)P(H)}{P(e)}$$

Posterior
How probable is our hypothesis given the observed evidence? (Not directly computable)

Marginal
How probable is the new evidence under all possible hypotheses (or: How likely have we observed the evidence regardless of hypotheses?)

$$P(e) = \sum P(e|H_i)P(H_i)$$

Bayesian Learning

- Let S_{-i}^{i} be the set of the opponent's strategies considered possible by player i, and H be the set of possible histories of the game
- Then we can use Bayes' rule to express the probability assigned by player i to the event in which the opponent is playing a particular strategy $s_{-i} \in S_{-i}^{i}$ given the observation of history $h \in H$, as

$$P_i(s_{-i}|h) = \frac{P_i(h|s_{-i})P_i(s_{-i})}{\sum_{s'_{-i} \in S_{-i}^{i}} P_i(h|s'_{-i})P_i(s'_{-i})}$$

Properties

- A very **intuitive** model of learning
- Players satisfying the assumptions of the model will have beliefs about the play of the other players that **converge to the truth**
- Players will in finite time **converge** to play that is **arbitrarily close to the Nash equilibrium**
- A player's beliefs will eventually converge to the truth if (1) it is using **Bayesian updating**, (2) is playing a **best response** strategy, and (3) the play predicted by the other players' real strategies is **absolutely continuous** with respect to that predicted by its beliefs
  - It will correctly predict the ***on-path portions*** of the other players' strategies
  - If an agent keeps playing its best response based on its beliefs of other agents, then eventually it will start to play the **best way** that it can be

Absolute Continuity

Every outcome (or action) that player 1 considers to have non-zero utility is also considered to have non-zero utility by player 2; the set of outcomes is the same

Definition 7.3.1 (Absolute continuity) Let X be a set and let $\mu, \mu' \in \prod(X)$ be probability distributions over X. Then the distribution μ is said to be absolutely continuous with respect to the distribution μ' iff for $x \subset X$ that is measurable it is the case that if $\mu(x) > 0$ then $\mu'(x) > 0$.

Properties 2

Even if a player models other players' true strategy incorrectly, the player can still be playing its best!

- The result does ***not*** state that players will learn the true strategy being played by their opponents
  - However, a player's beliefs must converge to the truth even when its strategy space is incorrect (does not include the opponent's actual strategy), as long as they satisfy the absolute continuity assumption
- If ***utility-maximizing*** players start with individual subjective beliefs with respect to which the true strategies are **absolutely continuous**, then in the long run, their behavior must be essentially the same as a behavior described by an **ε-Nash equilibrium**

If every player plays to maximize its utility, not just to be rational, and all have the same set of non-zero-utility outcomes (or actions), then eventually their play will converge to an equilibrium

Connection to MAS?

In MAS, an agent can observe actions of other agents, but not their intentions, and even more rarely their **types**

If actions correspond to evidence, then types correspond to hypotheses: If an agent knows the type of another agent, it can predict its actions generalizing from the characteristics of that type

Do we do that? What are some of the pitfalls?

Silly Question: If automobiles could only go forward and not backwards, what would happen?