Team

- Brenda Ng, LLNL
  - Machine learning; PI

- Kofi Boakye, LLNL
  - Signal processing; implementation lead

- Carol Meyers, LLNL
  - Applied math; theory lead

- Andrew Wang, MIT
  - Summer student intern; Renaissance man
Talk outline

- Goal
- Motivation
- Example applications
- Assumptions and strategies
- Single-agent decision process (POMDP)
- Interactive decision process (IPOMDP)
- Bayes-adaptive interactive decision process (BA-IPOMDP)
- Concluding remarks
Goal

- Advance modeling and response against human-like agents who seek to actively “game” against each other over the course of repeated interactions

- Build from current theory in artificial intelligence
  - Sequential decision-making frameworks

- “Bridge the gap” between theory and practice to solve real-world adversarial problems
Motivation

- Humans analyze many factors before acting
  - Current status
  - Opponent behavior
  - Past strategies (opponent and self)

- Drawbacks in traditional game theory (Nash equilibria)
  - No clear way to choose between multiple equilibria
  - Inability to deal with opponents that do not act according to equilibrium strategies

Can we develop computer systems that process decisions more like we do?
Assumptions and strategies

- Uncertainty about the (non-deterministic) environment
  - Maintain belief, or probability distribution, over states

- Example: card games
Assumptions and strategies

- Intelligent opponents (who also maintain beliefs about us)
  - Account for the opponent’s beliefs in *nested* models; more uncertainty inherent in more deeply nested beliefs
Assumptions and strategies

- Uncertainty about the effects of actions
  - Not entirely certain about how:
    - Environment state changes as a result of actions
    - Observations are related to environment state

- Treat transition model and observation model as part of the uncertain environment state

- Maintain beliefs over model parameters (in addition to the environment states)
To develop our model, we start with the single-agent decision process... the POMDP

- A *single-agent decision process* at each time step involves:
  - $s$: state of the environment, unknown to the agent
  - $a$: action that the agent performs
  - $r$: reward due to current state and current action
  - $z$: observation due to current state and previous action
Background: POMDP

- Common framework for planning in single-agent domains

\[ POMDP = \langle S, A, T, \Omega, O, R \rangle \]

- States \( S \)
- Actions \( A \)
- Transition function \( T : S \times A \rightarrow \Delta(S) \)
- Observations \( \Omega \)
- Observation function \( O : S \times A \rightarrow \Delta(\Omega) \)
- Reward function \( R : S \times A \rightarrow \mathbb{R} \)
Background: POMDP

- Common framework for planning in single-agent domains

Agent’s objective: optimize rewards given its beliefs
For adversarial modeling, we need an *interactive* decision process... the IPOMDP

- An *interactive decision process* involves (at least) two agents; their joint actions affect the next state.
- Each agent has its own *interactive states (is)*, with nested beliefs to predict the opponent’s action.
Background: IPOMDP

- Multi-agent extension of POMDP
- Supports decision-making in both cooperative and non-cooperative settings

\[ \text{IPOMDP}_{i,l} = \left< IS_{i,l}, A, T_i, \Omega_i, O_i, R_i \right> \]

- Interactive states \( IS_{i,l} = S \times M_{j,l-1} \) with \( IS_{i,0} = S \)
- Joint actions \( A = A_i \times A_j \)
- Transition function \( T_i : S \times A \rightarrow \Delta(S) \)
- Observations \( \Omega_i \)
- Observation function \( O_i : S \times A \rightarrow \Delta(\Omega_i) \)
- Reward function \( R_i : IS_i \times A \rightarrow \mathbb{R} \)
Background: IPOMDP

- Multi-agent extension of POMDP
- Supports decision-making in both cooperative and non-cooperative settings
To increase realism, we came up with an adaptive interactive decision process... the BA-IPOMDP

- A BA-IPOMDP allows uncertainty to be associated with the transition and observation functions via “augmented” Bayes-Adaptive interactive states (bais).
- A bais contains counts on previous state transitions and observations.
- The counts define the expected probabilities for T and O.
A number of computational challenges exist in solving a BA-IPOMDP

- Nested beliefs can lead to exponential increase in runtime for belief update
- Huge state space due to counts being part of the state
- Reachability trees with large branching factors
Simulation experiments: multi-agent tiger problem

- Two rooms/states: ferocious tiger in one room, jackpot in the other.
  - Tiger position resets when a door is opened.
- Three actions: \{open left door, open right door, listen\}.
- Six observations: \{growl from left side, growl from right side\}
  \times \{door creak from left side, door creak from right side, silent\}.
- Rewards: -100 for opening the tiger’s door, +10 for opening the pot of gold’s door, -1 for listening.
Results

- Learned values for observation probabilities converge to actual values.

- Learning agent earns more rewards than non-learning agent with incorrect assumptions.
Results

- Learning agents take more conservative actions, thus earn less rewards than non-learning agents.
Concluding remarks

- The POMDP and its extensions provide a natural way to model sequential decision-making under uncertainty.
- Major advances made in applying AI theory to real-world problems (mostly coordination between cooperative agents).
- In theory, proposed framework shows promise for modeling complicated human adversarial systems.
- In practice, deployment currently hindered by algorithmic complexity.

For technical details and references, please refer to our AAAI paper.