Introduction

• Theory of mind (ToM): humans’ ability to infer and understand the beliefs, desires, and intentions of others [4].
• Cognitive Machine Theory of Mind (CogToM): a framework that relies on a general cognitive theory of decisions from experience, Instance-Based Learning Theory (IBLT) [3].

Instance-Based Learning Theory

• IBLT explains human learning in dynamic decision problems [3].
• An “instance,” a memory unit in IBLT, is represented by three elements: a situation (s), a decision (D) (or action s taken in state s), and a utility (U) (expected utility or experienced outcome of the taken action taken in state).
• IBLT uses the Activation equation of the ACT-R architecture [1] for representing how readily available the information is in memory.

CogToM: A Cognitive Machine Theory of Mind Framework

• An observer is a cognitive model built on IBLT [3] that builds ToM by observing the actions of agents playing in a gridworld.
• The IBL observer can predict the agent’s future behavior, such as a next-step action or the agent’s desired target in a new gridworld.

• A gridworld is a sequential decision-making problem wherein an agent moves through a $S \times S$ grid ($S = 15$) by making decisions (i.e., up, down, left, right) to search for targets.

Models of Acting Agents in the Gridworld

• Random agent: agent $A_k$ selects an action $a$ in state $s$ based on the probability $p(a|s)\sim \text{Dir}(a)$.

• Reinforcement Learning (RL) agent adopts a tabular form of Q-learning algorithm, a well-known temporal difference approach [6].

• Instance-based Learning (IBL) agent uses the memory and learning mechanism in IBLT. It selects the action with the highest expected utility using the blended value.

IBL Observer

• Derived from the observable actions of the agent, the IBL observer infers the agent’s true reward function.
• Based on the inferred reward, the IBL observer makes the prediction about the agent’s behavior in the new environment.
• The “past experience” of the IBL observer is implemented by inserting “pre-populated instances” in the model’s memory.

Experiments

Following [5], three experiments were conducted: (1) an arbitrary goal task, (2) a goal-directed task, and (3) a false belief test of ToM.

Experiment 1: Arbitrary Goal with Random Agents

• Agents’ goal: obtain one of the four colored objects within 51 steps.
• IBL observer: predict the initial action of the random agents in a new gridworld, given the agents’ trajectories in a past gridworld.

Experimental Setup

• Different types of random agents: $\alpha \sim \{0.01, 0.03, 0.1, 1\}$.
• Different number of past gridworlds: $N_{\text{past}} \sim \{0, 1, 5\}$.
• Number of observed agents for each type is 100.
• Evaluation metric: the accuracy of accurately predicted actions relative to the agent’s true next action.

Results

- $N_{\text{past}} = 0$: the observer’s prediction is independent of $\alpha$.
- $N_{\text{past}} = 1$ and 5: the IBL observer’s accuracy increases.
- Accuracy diminishes as $\alpha$ increases: it is easier for the IBL observer to predict the agents’ behavior with near deterministic policies.

Experiment 2: Goal-Directed Task with RL Agents

• Agents’ goal: obtain a particular object that has the highest reward within 51 steps. Consuming any of the other objects leads to the termination of the episode.

• IBL observer’s goal: learn to infer which object the RL agent desires to consume, and then predict (1) the next-step action that the agent would take, and (2) the object the agent would consume in the new environment, given either full or partial observation of the agent’s trajectory in a training gridworld.

Experimental Setup

• Each agent $A_k$ is driven by a fixed reward, $r_{k,j} \in \{0, 1\}$, for consuming an object $a_j$ where $j = 1, \ldots, 4$.
• For the analysis of partial trajectories, $N_{\text{past}} = 1, 10$.
• Number of RL agents is 100.
• Evaluation metric: the difference between the RL agent’s true behavior (the ground truth) and the IBL observer’s predictions.

Results

- Prediction accuracy: (1) next-step action is $0.515 \pm 0.08$; and (2) goal consumption is $0.687 \pm 0.09$ with $95\%$ confidence level.
- Regarding partial trajectories, the IBL observer’s prediction accuracy is improved when increasing $N_{\text{past}}$.

Experiment 3: False-belief Test with Three Agents

- Sally-Anne test is mapped onto the gridworld setting as follows:

1. Sally-Anne test
2. Gridworld task

Experiment 3: False-belief Test with Three Agents

- Sally-Anne test: an agent $A_j$ is trained to be a blue-object-prefering agent.
- $A_j$ is forced to reach a subgoal where it will see the preferred object, but not the swap.
- After reaching the subgoal, agent $A_j$ will continue towards the preferred object.
- The IBL observer observes the agent consuming an object.

Results

- Use the IBL process of IBLT [3] and the formulations of the ACT-R architecture [1] for memory-based inference to demonstrate how ToM develops from observation of other acting agents’ actions.
- Illustrate the ability of the IBL observer to predict next-step action, intention, and false beliefs in novel situations.

Conclusions

- Evaluate (1) how the agent behaves in the swap and no swap settings and (2) how the IBL observer performs when observing different types of agents in the two settings.
- Evaluation metric: Jensen-Shannon divergence ($D_J$) between the probability distribution over the locations of the four objects that the agent consumed in the swap and no swap events.

Acknowledgements

This research is based upon work supported by the Defense Advanced Research Projects Agency (DARPA), award number: F306022636.