Robots and humans teaming up in pursuit-evasion games: modeling and learning issues

Shen Li (shenl1), Luqing Zhang (luqingz), Siliang Lu (siliang1)

November 8, 2016

1 Pursuit-evasion Game

1.1 Goal

- Pursuit-evasion is a classic problem for mathematics and computer science where we use to study the collaboration between human and a multi-agent system. We will implement this problem in a game setting where 1 human player controls 1 pursuer and there are several other robot pursuers collaborating with the human pursuer. The goal for these pursuers is to catch all the robot evaders who try to escape through the exits.
- World = undirected graph with exits
- Pursuers and evaders take turns to act
  - Pursuers move at the same time triggered by the human player
    - Human pursuer has limited perception and decision, which means that human has an imprecise model of the world.
    - The robot pursuer updates its models for the human partner and all the evaders. Then it will predict the world state transitions based on these models and act optimally to maximize the rewards. Its *optimally* take account for human *non-optimality* in decision-making when being in the same team.
    - Human pursuer and robot pursuers are sharing a MDP model so they are making decision together. They cooperate together closely while the evaders never do, which gives pursuers chance to win.
  - Robot evaders move at the same time optimally to maximize their rewards.
- Robot evaders will be removed when they cannot move except stepping onto a pursuer or an evader.

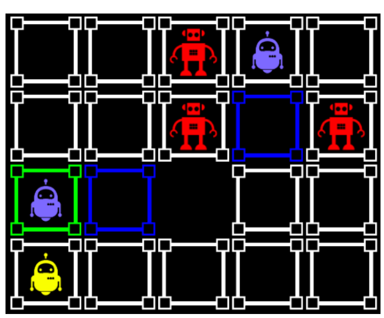

Figure 1: The pursuit-evasion game (agents can be teleported through swirls)

1.2 Related Work

- The robot models human stubbornness based on human behaviors in a bounded memory of robot and decide whether to submit to human or not in a table carrying task ([1]).
- The robot has an MDP to make its own decisions and an MDP for modeling the behaviors of its human partner. Then it can cross train both of the models with its human partner ([2]).

1.3 Motivation/Contribution

- Usually people have to use reinforcement learning to train a robot model used for this particular person, which requires lots of data before start. We want the game to play at start. We need to learn the parameters on-line quickly to get a team policy.
- We propose a typical way to solve a human-swarm interaction problem
  ⇒ We have a problem.
  ⇒ What model we use?
  ⇒ How to take into account the fact that human acts optimally based on limit information?
  Human less capable ⇔ more *random* in the transition function of human action
  ⇒ What parameters we want to use to shape the model in real time based on the human input?
  ⇒ We can make the robot pursuers act optimally based on its model of the human player and all the opponents.
  ⇒ This is very expensive computationally so we have to do approximation, which will make robot pursuers less optimal
  ⇒ How to deal with robot pursuers sub-optimality vs human sub-optimality?
- A model of human capability in this game based on 3 human abilities (we can use reinforcement learning to train the parameters)
  - ability to discriminate among choices = difficulty of the decision ⇐ $||a_h||$
  - ability to look ahead = ability to foresee the impacts of decisions ⇐ min(number of turns until a capture)
  - ability to grasp useful knowledge about the current state = $Q_h(s, a)$
  - ability to understand or predict the robot teammates' reasonings so that he can act cooperatively = look back in the past

1.4 *Admissible* Plan :)

- By 11/20: game able to play with naive human model
- By 11/30: game able to play with a good human model based on literatures
- By 12/10: game able to play with a good human model based on literatures which is trained by reinforcement learning
- (may delay because of the computational difficulty in the big MDP)
- shenl1 and siliang1 - model robot and human, luqingz - game design, together - reinforcement learning

References

[1] Stefanos Nikolaidis, Anton Kuznetsov, David Hsu, and Siddhartha Srinivasa. Formalizing human-robot mutual adaptation via a bounded memory based model. In *Human-Robot Interaction*, March 2016.

[2] Stefanos Nikolaidis and Julie Shah. Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 33–40. IEEE Press, 2013.