Quiz 2

ASEN 6519-001: Advanced Survey of Sequential Decision Making

November 17, 2021

Question 1. (1 pt) Suppose that SARSOP is being used to solve a two-state POMDP. At some point in the computation, suppose there are two alpha vectors, $\Gamma = \{[1, 0], [0, 1]\}$, and two belief points in the tree, $B = \{[0.3, 0.7], [0.2, 0.8]\}$. Will SARSOP prune any α-vectors in this case?

Question 2. Answer the following questions about the DESPOT family of algorithms

a) (1 pt) Suppose that you use the DESPOT algorithm with K=6 scenarios to solve a POMDP with 5 actions, 8 possible observations, and 3 possible states. What is the maximum number of children that any node in the tree could have? Explain.

b) (1 pt) Suppose that you have used reinforcement learning to learn an approximate policy for a POMDP. Is it possible to use this policy to generate upper or lower bounds for use in AR-DESPOT? Explain.

Question 3. (1 pt) POMCPOW and DESPOT-α both use weighted particle collections to represent beliefs within the tree. However DESPOT-α additionally uses α-vectors to share value estimates between nodes. Why can't POMCPOW use this α-vector sharing trick?

Question 4. Consider a "Plus-Minus" game with the following rules: Player 1 chooses either +1 or -1 and then Player 2 chooses either +1 or -1. If the *absolute value* of the sum of both players actions is greater than 1 (i.e. $|a_1 + a_2| > 1$; the signs of the actions are the same) then Player 2 wins. Otherwise Player 1 wins.

a) (1 pt) Suppose that Player 2 gets to see Player 1's action before playing (i.e. it is a complete-information sequential game). Draw the game tree. Which player has an advantage?

b) (1 pt) Now suppose that Player 2 plays without seeing Player 1's action (i.e. it is an incomplete-information game). Indicate which nodes of the tree from part (a) belong to the same information set. How does this change the Players' strategies?

Question 5. (2 pt) List 3 algorithms or systems discussed in the readings or lecture that are capable of playing chess and list one attribute that makes each distinct from its predecessors.

Question 6. (1 pt) Heads-Up No Limit (HUNL) Texas Hold 'em Poker, the game addressed by Libratus, involves partial observability over discrete states, actions, and observations. Suppose that, based on those problem characteristics, another student asks you if it is a good idea to use POMCPOW to solve HUNL Texas Hold 'em Poker. How would you respond?

Question 7. (1 pt) Suppose that you create a reinforcement learning agent with the same neural network architecture as the AlphaStar agent. You begin with supervised training on human data just like AlphaStar, and then train it further by repeatedly playing the built-in elite bot. Call this the "bot-trained agent". How would you expect the bot-trained agent to fare in league play against humans? Explain how the AlphaStar agent differs from this bot-trained agent.