Assignment 6: Joint Inference of Belief and Desire

1 Introduction

As the final project of the class, you will implement a Bayesian model to infer beliefs and desires about an agent given an observed position trajectory sequence. This requires integration of knowledge and skills from the previous weeks including value iteration, implementing a POMDP, and sequential Bayesian Inference.

We replicate the results of Baker, Saxe, Tenenbaum (2011) and Baker, Jara-Ettinger, Saxe, Tenenbaum (2017) with some simplifying assumptions. We will draw from their concrete example of a hungry graduate student looking for lunch at one of the food trucks on campus:

There are three food trucks that come to campus, each serving either Korean food, Lebanese food, or Mexican food. The university provides only two parking spots, so at most two trucks can be on campus on any given day; parking spots will never be empty and there is only one truck for each cuisine type. When the student leaves her office, she can see the truck parked in the near spot in the southwest corner of campus. The other truck is parked in the northwest corner of campus is around a set of buildings and is not visible unless the student walks around the corner to see what is there.

Based on the agent's positional trajectory, an observer can model at different time points, what the student believes to be true about the world and the student's relative food preferences (desires). You will be given a couple of state trajectories in possible worlds and asked to model a theory of mind inference about the beliefs and desires of the agentas an observed.

The principles governing the relation between the world and the agents beliefs, desires, and actions can be naturally expressed as a POMDP. An agent's observations of the truck in either location allow the agent to update its beliefs to be consistent with only the worlds possible given that observation. The observer maintains a hypothesis space of beliefs and desires the agent could have and evaluates the likelihood of generating the observed behavior from those beliefs and desires. To construct the hypothesis space the observer can model how the agent should act for each desire and belief using the POMDP framework. Then an observed state trajectory and world knowledge are used to evaluate what beliefs the agent has and how likely each desire is for different time points.

2 Inputs

You are given:

- State trajectories: a list of positions an agent is observed to be in for a time sequence
- World: the world in which the observer saw the agent take those steps
- Environment: The state space and action space
- Reward Values: the numeric costs and rewards for different states

You may assume that the agent always starts out with no prior knowledge of the worlds at time 0.

3 Setup

Implementation details are up to you, but the belief transition, reward belief table, policy over beliefs, and final inference should follow these principles.

Environment and Observations

Here is the two dimensional visualization of a simplified food truck environment we will use. The trucks are located in position (0,0) and (4,3). The shading on the image correspond to which positions certain trucks are visible in (where the agent will receive observations). Observations are recieved with probability 1 and always correct. There is no chance the agent will be mistaken about which truck is in the location once she is in a position to see it.

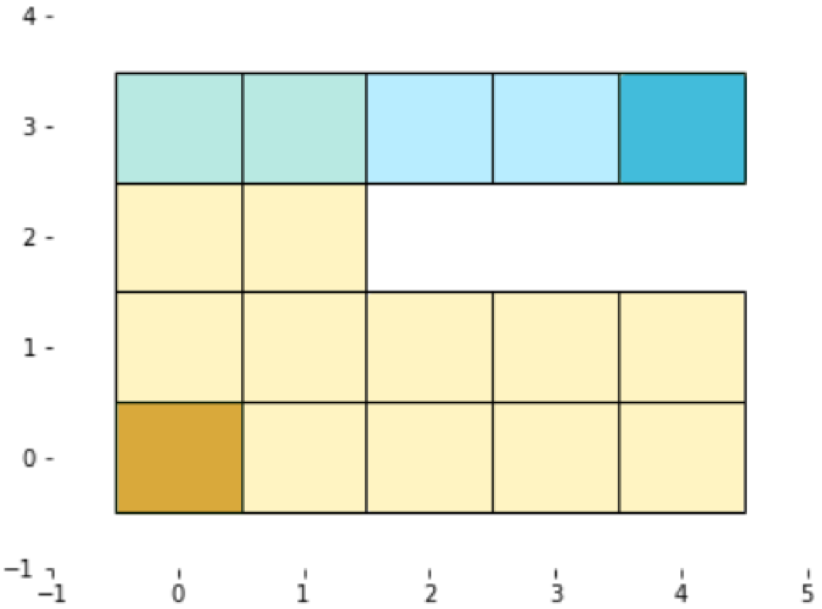

Figure 1: POMDP Environment: Truck 1 Location in mustard and positions where truck 1 is visible in yellow. Truck 2 location in blue and positions where truck 1 is visible are lighter blue

The agent knows its position on the grid but there is uncertainty about which trucks are parked in which spots until after observations are received. There are six possible worlds: KL, KM, LK, LM, MK, ML. The first letter corresponds to the truck in position (0,0) and the second to the truck in (4,3).

State Representation:

States are now tuples of form: ((position), (belief state)), where position is (x,y) the coordinate on the 2-D environment grid, and the belief state is a tuple of length 6 where each position is the probability the agent is that world. The worlds are: KL, KM, LK, LM, MK, ML, in that order. For example, if the agent knows it is in world KL with probability 1, its belief state will be (1,0,0,0,0,0,0). If it knows the tuck in position (0,0) is a L, its belief state is (0,0,.5,.5, 0,0), etc. You should also account for the belief state with no knowledge about the world (1/6,1/6,1/6,1/6,1/6). This belief state will only be possible at time 0 as upon the first action, no matter what it is and no matter what position the agent is, it will receive an observation. You can round and represent this belief as (.17,.17,.17,.17,.17,.17) to avoid dictionary key errors.

Desires and rewards:

Desires are equivalent to preference rankings. There are six possible ways to rank the three types of food. The type of the food that is most preferred will be given a reward of 100, the type of food that is in the middle should be constructed to have a reward of 75 and the least preferred food has a reward of

50. All actions should have a cost of -1 except stay (0,0), which has the reward of -.1. You are in charge of constructing your reward tables and belief rewards from this information. The final belief reward will be of form { state : { action { next state: ρ(state, action) }}}

Transition:

The position transition of the agent is deterministic (the intended action is always taken with probability 1 if possible in the environment, otherwise the agent stays in the same position). The belief transition (unlike before) is not conditioned on which particular observation is received, but rather is able to plan for any possible observation that might occur. You should assume all observations are equally probable. This means that your belief transition should map to multiple next possible beliefs when new information is received, one belief for each possible instance the observed truck could take. That is, if you had no information before and then receive the observation of the first truck your set of next possible beliefs should include the three belief states where the first truck is known but the second is not.

Policy:

In contrast to the previous assignment, the POMDP policy will be generated once for each desire using the belief transition and belief reward generated for that preference set, no matter which world the agent is in. That is, the policy should be a mapping from position and belief state to action for any truck combination that it is possible to observe. This will result in 6 policies that together can take any possible belief state, position, and preference set and return the actions to take.

Inference: Inferences about the agent's beliefs are deterministic, because the observations the agent receives always occur within the visible zone and are always correct. The Bayesian inference to infer desire follows that of Assignment 3. You should assume that the prior over all desires is uniform.

4 Writeup and Output Requirements:

You should upload your code as a python file or Jupyter Notebook. In addition, you should include a PDF write-up.

The write-up should include:

1. Plot of posterior probabilities over desires across time for each trajectory and world given. Time on the x-axis, a normalized probability on the y-axis, and a different colored line for each of the 6 possible desires with a labeled legend.
2. The agents' inferred beliefs across time (as a visualization, table, or description) for each trajectory and world example
3. A visualization of the calculated policies. As there are so many, please only include the visualization for the policy that corresponds to the preference set $K > L > M$ with belief states (0,0,0,0,0,1), (.5,.5,0,0,0,0), (0,0,.5,.5,0,0), (0,.5,0,.5,0,0), (.17,.17,.17,.17,.17,.17)
4. What happens to the policy when you change gamma (the discounting factor)? Provide an intuitive explanation for why this could be the case.