Karim Soliman
CSEP 599: Robotics
June 10, 2018

Lab 14: Human in the loop

Cozmo Picking up trash through TAMER algorithm

What would we do?

We will start by discussing what humans would do to solve this problem. First we would look at the table and identify all the objects determine what is trash and what isnt, identify all the empty space, and determine the location of the trash bin (the goal). We would then move to the trash, pick up the trash, go to the trash bin, and drop the trash. We would repeat this until all the trash is cleared.

How does this apply to cozmo?

To take this approach in cozmo terms we will split the problem into 4 stages not counting an initial set up stage. The set up stage would be gathering the info on what is on the map. We could be provided with a map of all the trash or we could use the robot camera to identify all the trash. For our purposes we assume we are given a map with all the trash locations as well as the location of the trash bin. We also know of any non trash items as well as the edges of the table. We will map this into a grid where each cell is empty, has trash, has non trash, or has goal. We will assume that the trash bin is a part of the table (maybe its hanging on the edge or is an empty whole in the middle of the table).

Now that we have the state map, we need to define the actions the robot can take. At any point cozmo can move one cell forward, back, left, or right. We will assume turn and go right/left is one movement for cozmo. When an adjacent cell is a trash item, we have the action of lift, and when the adjacent cell is trash bin, we have the action of drop. Not all actions will be available in all stages. So lets get into the stage, what they are, and what actions can happen.

The Stages

First stage is navigation to the trash. The actions available in this state are forward, backward, left, and right movement. When the robot moves towards a trash item the human would give a positive reward as defined in the TAMER algorithm. When the robot moves closer to non trash item or isn't heading towards a trash the user will give a negative reward. The one thing that is a little problematic is the difficulty of the human knowing if the robot is heading in the right direction. There may be a trash item behind a non trash item. Would the human administer a negative reward even if robot was heading to trash by going towards then around the non trash item? I think that over time issues like this will be worked around due to the repeating reward/negative reward structure. Even if once or twice the human gives the wrong reward, over time the robot will know the appropriate movement. The result of the learning in this stage is the robot will move towards the trash and avoid non trash items.

The second stage is picking up the trash. This is a state where we have an adjacent trash item. In this state another action that is now available is lift trash. This is in addition to the move left, right, forward, backward. The trash would still be counted as an obstacle so moving in the direction of the trash item wouldn't happen. When the robot would pick up the trash the user would give a positive reward, and a negative reward otherwise. This will result in a very simple action, if next to trash then pick it up, there should be no ambiguity on what action is best.

Stage 3 will be navigating towards the goal. The only actions allowed here are the movement actions of left, right, forwards, and back as in stage 1. The user would administer a positive reward when approaching the trash bin and a negative reward otherwise. Here we run into the same issue of inconsistency in reward giving when the user is not sure if the robot is heading to trash bin or an item infant of the trash bin. This is overcome through the many learning iterations. The result of this step is the robot has a trash item and is standing near the trash. When there is a trash item with the robot and the robot is near a trash bin we move to stage 4.

Stage 4 is complimentary to stage 2 where the robot will just have to drop the trash item. In this state the action available is drop as well as the movement actions left, right, forward, and backwards. The user would give a positive reward when the robot drops the trash item and a

negative reward otherwise. This will result in the simple action of, if next to trash bin and have trash then drop it.

Thats It!

The learning and reward given by user as described above (modeled on the TAMER algorithm) will result in the robot learning to navigate to trash, pick it up, navigate to trash bin, and drop the trash. The robot would do this without dropping off the table (since the map provides the empty spaces the robot is allowed) and without trashing any non trash items since the user would have given negative reward for approaching and picking up non trash items. The robot simulation would go through, try an action, and get reward. The model would then be learned based on the reward and all future actions would be based on the previous rewards.