Guiding Robot Project

Initial formation by Anushka, Hadar and Gil

February 2024

1 Problem requirements

Is the simulation the objective or a way to achieve it?

Our objective is to develop a simulated system to control a robot that is tasked with safely guiding a visually impaired person around obstacles to the destination.

this is choice - not a requirement

The method this project requires is Offline Reinforcement learning, using human-human interactions as learning data, similarly to work done by Hong et. al. (reference).

World assumptions:

1. The guiding agent ("Human player") and human agent have a tether connecting them.
2. When the tether is at max length, human feels "pull" signal (arrow in the direction of the tether).
3. Once the tether is at max length, the robot cannot pull the human, but the human can pull the robot (causing it to move in the tether direction).

2 Formulation

2.1 States and Observations

Our map is a 2D obstacle map as shown in figure 1.

A **state** is an nxn occupancy matrix which makes a grid representation of the world, and specify what is in each tile of the grid (similar to "frozen lake" from intro to AI and "multi taxi" from CLAI course):

- Blank space [-]
- Obstacle [X]
- Human agent [H]
- Robot agent [R]
- Goal region [G]

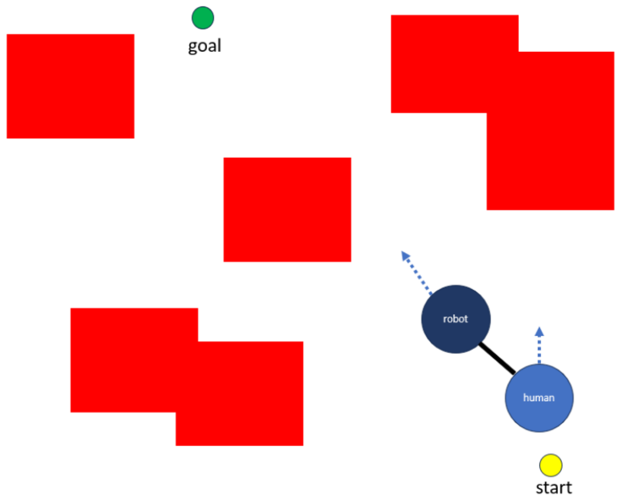

Figure 1: 2D Map example

The robot's **observation** is comprised of:

- The state (occupancy matrix)
- Tether length left - which is fully defined by the human and robot locations, along with the maximum tether length (parameter)

So the robot has full observability of the human?

At this point, we consider the entire map to be fully visible to the robot. In the future we might consider an "exploration" scenario, in which the robot can only see up to a finite-horizon at each step, with the rest of the map being "unknown", so that the robot must explore the map to find the goal and obstacles (to make sure the problem is Markovian, once the robot "sees" a tile once, it "remembers" what's there for the rest of the episode, similar to strategy games).

with the person ?

2.2 Actions

The agent actions involve moving around the 2D map. We discussed 4 levels of complexity for the actions:

1. Basic Manhattan grid - Actions are {up, down, left, right, noop}
2. Heading angle with fixed step size - Actions are $\left\{360\frac{n}{N}\right\}_{n=1}^{N}$
3. Heading angle and step size - Actions are $\left(r, \left\{360\frac{n}{N}\right\}_{n=1}^{N}\right)$
4. Heading angle and velocity/acceleration - accounting for dynamics and inertia

Great - I am assuming here that this is all the robot's movements and that it has perfect localization capabilities. Right?

While higher level of complexity better models the real-world problem, at this point, we choose not to explore the last 2 levels, as they could complicate human-human data collection.

2.3 State-Action Transition

I am missing a formal definition

While the kinematics of the problem are deterministic, we note two important elements that are variant:

1. The human position is independent of the robot actions.
2. The tether constraints the robot's movement - so that even if the robot chooses to move in a "legal" direction (not into an obstacle or out of the map), in the consecutive state it might end up "moving" in the opposite direction if the human drags it.

2.4 Reward

We account for several components for a state-action-state' reward:

- Human collision (s) - High negative reward for human collision with obstacle.
- Robot collision (s) - Small negative reward for robot collision with obstacle.
- Arriving at goal (s) - High positive reward for arriving at the goal region.
- Number of steps (s) - Small negative reward which accumulates for each step the human does not reach the goal.
- Optional - Influence reward (s,a,s') - Positive reward for human moving in the direction of the robot (= robot "succeeds" in influencing the human)

nice! (but again - this is lacking formal definitions)

2.5 Problem parameters

these are simplifying assumptions

1. Grid size - If the map is too big, rollouts could be too long to help effective learning. seems weird that you can control the size of the deployment grid - you can state your limitations.
2. Maximum tether length - A too long tether could lead the the robot solving reacing the goal without being able to guide the human to it. A tether too short could lead to a somewhat degenerated problem where the human and robot practically move together.

3 Modifications for human-human data collection

The data which will be used in the Offline-RL algorithm will be collected using a human-human interactive game.

two?

The game will have to human players - one acting as the guiding agent and one acting as the blind human. The guiding agent has full visibility of the map and full control of its motion. The human agent sees nothing, except for when the tether is at full length, then the human sees an arrow pointing in the direction of the tether.

Both agents see:

- A ”Success” screen once the human reaches the goal area
- A ”Game Over” screen if the human collides
- A step counter showing how many steps the human made (OR time elapsed)

4 Evaluation of the agent (after implementation)

The agent evaluation will be performed via live experiments, using the same interface as the human-human data collection game, only the guiding agent will be acted out by the evaluated agent. The baseline could be a ”naive” agent which is unaware of the human and aims to reach the goal by itself.