Intro to AI Project 3 Write Up

April 10, 2021

On my honor, I have neither received nor given any unauthorized assistance on this assignment. All of our work and ideas are our own. We did not use any other external sources.
- Michael Zhang, mbz27, Section 1

On my honor, I have neither received nor given any unauthorized assistance on this assignment. All of our work and ideas are our own. We did not use any other external sources.
- Prasanth Balaji, pmb162, Section 3

Group members: Michael Zhang, Prasanth Balaji

Problem 1

X_i = target in cell i
Y_j = search result in cell j

$P(X_i|Y_j = 0)$: Case $i = j$

$\Rightarrow$ Use Bayesian Method

$$\Rightarrow \frac{P(Y_j = 0 \mid X_i = 1)P(X_i = 1)}{P(Y_j = 0)}$$

$\Rightarrow$ Marginalize denominator

$$\Rightarrow \frac{P(Y_j = 0 \mid X_i = 1)P(X_i = 1)}{P(Y_j = 0 \mid X_i = 0)P(X_i = 0) + P(Y_j = 0 \mid X_i = 1)P(X_i = 1)}$$

$P(X_i|Y_j = 0)$: Case $i \neq j$

$$\Rightarrow \frac{P(Y_j = 0 \mid X_i = 1)P(X_i = 1)}{P(Y_j = 0 \mid X_i = 0)P(X_i = 0) + P(Y_j = 0 \mid X_i = 1)P(X_i = 1)}$$

$$\Rightarrow \frac{P(X_i = 1)}{P(Y_j = 0 \mid X_i = 0)P(X_i = 0) + P(Y_j = 0 \mid X_i = 1)P(X_i = 1)}$$

The probability of the target in cell i given the observations at time t and a failure in cell j can be expanded in the following way using Bayes' theorem. There are two distinct cases, one in which i equals j which means the same cell is being checked for the target given a failure. The other case is when i does not equal j, which means two different cells are being compared. In both cases, the numerator is expanded to P$(Yj = 0 \mid Xi = 1)$ * P$(Xi = 1)$. P$(Yj = 0 \mid Xi = 1)$ represents the likelihood, which represents a failure in cell j given that the target is in cell i. In the case in which i equals j, this will be represented by the false negative rate which is based on the terrain of the cell. As the project description mentions, "the more difficult the terrain, the harder it is to search - the more likely it is that even if the target is there, you may not find it (a false negative)". On the other hand when i does not equal j, this will always equate to 1 as there is a 100 percent chance of the target not being found in cell j given that it is in cell i. P$(Xi = 1)$ simply represents the probability of the target being in cell i and is known as the prior probability. The denominator P$(Yj = 0)$ which represents the target is not in cell j, can be marginalized so that it accounts for all possible scenarios of the target not being found in that specific cell. The marginalization of P($Yj = 0$) breaks up the probability into terms that we already know: P($Yj = 0 \mid Xi = 0$), P($Xi = 0$) , P($Yj = 0 \mid Xi = 1$), and P($Xi = 1$) . Using all this information, the probability of the target in cell i given the observations at time t and a failure in cell j can be found and is essential in updating our belief state for each cell. Now we can analyze cells with a higher probability based on this information.

Problem 2

X_i = *target in cell i*
Y_i = *search result in cell i*

$$P(Target\ found\ in\ Cell_i \mid Observation_t)$$
$$\Rightarrow P(Y_i = 1 \wedge X_i = 1 \mid Observation_t)$$
$$\Rightarrow P(Y_i = 1 \mid Observation_t)\, P(X_i = 1 \mid Observation_t)$$
$$\Rightarrow P(Y_i = 1)\, P(X_i = 1 \mid Observation_t)$$

The probability of the target being found in cell i given the observations at time t essentially represents a success given the belief obtained so far. The probability of the target being found at i is equal to the probability of the target being at i and receiving a success at i based on the information, which both represent the scenario of the target being found. Now we can distribute the " | *observation t*" to both cases of the scenario. The first part can essentially be P($Yi = 1$) , because Yi is independent of the observation at time t. The second part will be P($Xi = 1 \mid Observations\ t$). As a result, the P($Target\ is\ found\ in\ Cell\ i \mid Observations\ t$) is equal to P($Yi = 1$) * P($Xi = 1 \mid Observations\ t$).

Problem 3

On average we saw that agent 2 performed slightly better than agent 1. This is probably due to the fact that agent 1 iteratively travels to the cell with the highest probability containing the target and searches that cell. However, the highest probability of containing a target doesn't necessarily guarantee success. If it fails to find the target due to false negatives as a result of the terrain, then it will take a much longer time to come back to that specific spot to detect a success as compared to agent 2. Agent 2 considers the probability of finding the target when searching the tile as well as the belief of it actually being there. Therefore, agent 2 should be superior to agent 1 and the this is evident through the statistics.

Average Score (Manhattan Distance + Number of Searches)
after running 500 times:

Agent 1: 197.48, Agent 2: 183.62

Problem 4

When running agent 1 or agent 2, the most significant portion of their final score was mainly due to the Manhattan distance that the agents had to cover to get to the desired cell that contained the highest probability of belief or the highest probability of finding the target. Since both agents need to traverse this distance regardless, we thought it would be beneficial for the agents to pick an additional cell to search along the path. Since this wouldn't add too much to the final score and will result in more cells searched. In addition, by choosing one more cell to search, we can somewhat make sure that the original cell we were traversing to was actually the right cell to move towards since this will give us one more data point to add to our probabilistic calculations. We made two improved agents, one that will pick the highest probability of finding the target along the path (improved 2) and one that will pick the highest probable belief along the path (improved 1). Surprisingly even though our previous results indicated that agent 1 usually resulted in a poorer score than that of agent 2 in our improved agents, it was the opposite. Improved agent 1 (that used the probable belief to choose the additional cell to search) had a better score than improved agent 2 (that used the probability of finding the target to choose the additional cell to search). This is probably because when agent 1 searches a cell containing the target and does not find it, the agent will take a longer time to search the cell again compared to agent 2. But since we are checking more cells in the improved agent, it results in a short time to search the same cell again.

Average Score (Manhattan Distance + Number of Searches)
after running 500 times:

Improved Agent 1: 135.00, Improved Agent 2: 133.09

Documentation

How to run the project:

Each iteration will run agent 1, agent 2, improved agent 1, and improved agent 2 and will record the scores and average them based on the number of iterations
Please enter the arguments in this order [dimension greater than 1] [iterations greater than 0]

Command:

- java SearchDestroy [dimension greater than 1] [iterations greater than 0]