Text

Description automatically generated

Graphical user interface, text

Description automatically generated

S (state spaces) are a set of positions in this 11 by 11 matrix (from point (0, 0) to (10, 10)) except the walls.

A (action spaces) are actions selectable from each location, including left, right, up and down.

A picture containing text, electronics

Description automatically generatedGraphical user interface, text

Description automatically generated

From the graph showing in the left, there are 44 blue blocks, 40 green blocks, 15 orange blocks and 4 red blocks. They represent number of directions that each state could move in that block to avoid being hit by the walls. (B=4 directions, G=3, O=2 and R=2). (10, 10), the star position is not being considered as it arrives at the final position then the agent would go back to the original point.

as non-zero rows could be considered as the possible actions from each s to s’ as r is always 0 expect the reward is 1 for agent arrive at the final position.

Therefore, for all s in B region, possible actions could be 4\*3(4 directions \* 3 selections/direction), 44\*4\*3= 528. For all s in G region, possible actions could be 4\*3(including stay when hitting the walls), 40\*3\*4 = 480. For all s in O region, possible action could be 2\*3(not hitting walls)+2\*2(hitting walls), 15\*(2\*3+2\*2) = 150. For all s in R region, possible action could be 4\*3, 4\*4\*3=48. So the total estimation for the conditional probability table is 528+480+150+48=1206. Approximately, 1200 non-zero rows exist.

A screenshot of a computer

Description automatically generated with medium confidence

Text, letter

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence



Without using to implement the discount, the maximum return is always +1 whatever time step from the maze. To effectively communicate to the agent, we need to add punishment or cost with each time step before escape.

A picture containing text

Description automatically generated

Chart, scatter chart

Description automatically generated



Table

Description automatically generated

Text, letter

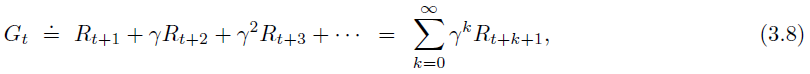
Description automatically generated

Text

Description automatically generated

Diagram

Description automatically generated



Text, letter

Description automatically generated

Text

Description automatically generated

Reward is important in an episodic task. By taking negative reward, the agent would finish the task faster. Thus, adding a constant C would change the task, especially when the constant changes the sign of the equation. If the negative reward is still negative but smaller, it may change the direction of an agent. But in some circumstance, the task may be unchanged.

Text

Description automatically generated

Chart

Description automatically generated with medium confidence Diagram

Description automatically generated



Text

Description automatically generated

Table

Description automatically generated with low confidence



Text

Description automatically generated

A picture containing text

Description automatically generated

Text, letter

Description automatically generated

A picture containing text

Description automatically generated

Text

Description automatically generated

A picture containing text

Description automatically generated

The value function is as n is the number of state (e.g. L, A, B, …, R) and i is position of the state.

A picture containing text

Description automatically generated

Diagram

Description automatically generated

A picture containing text

Description automatically generated

Chart, Word

Description automatically generated

Text, letter

Description automatically generated

A picture containing text

Description automatically generated

Text, letter

Description automatically generated

A picture containing text

Description automatically generated

Text, letter

Description automatically generated

According to equation (1) and (2), we could get .

In order to get the maximum value of , , and .

Diagram

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated