

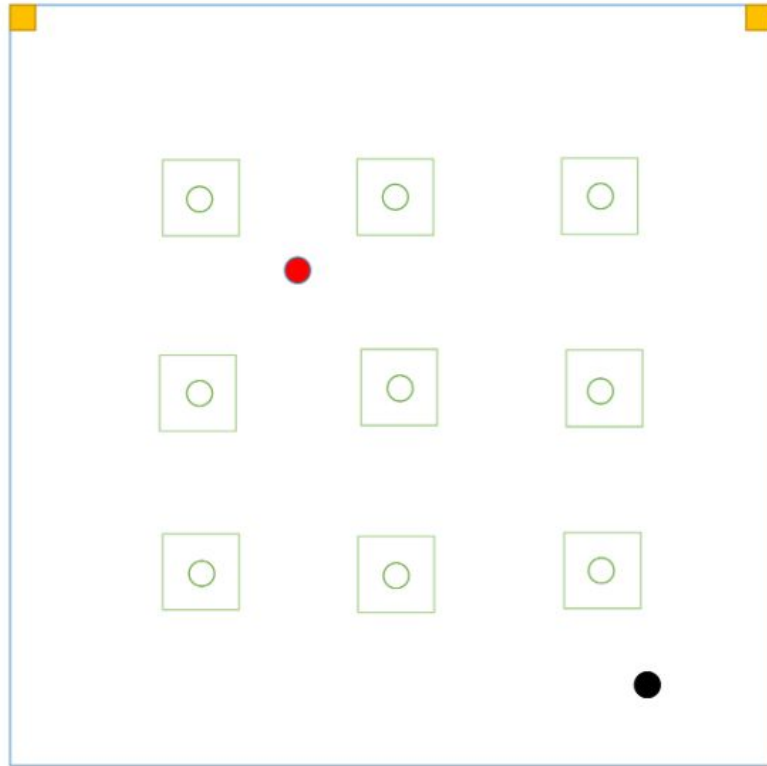
Hide and Seek Using Function-based Q learning

Yu-Hui Chen
Wen-Hao Kao
Mu-Lin Tsai

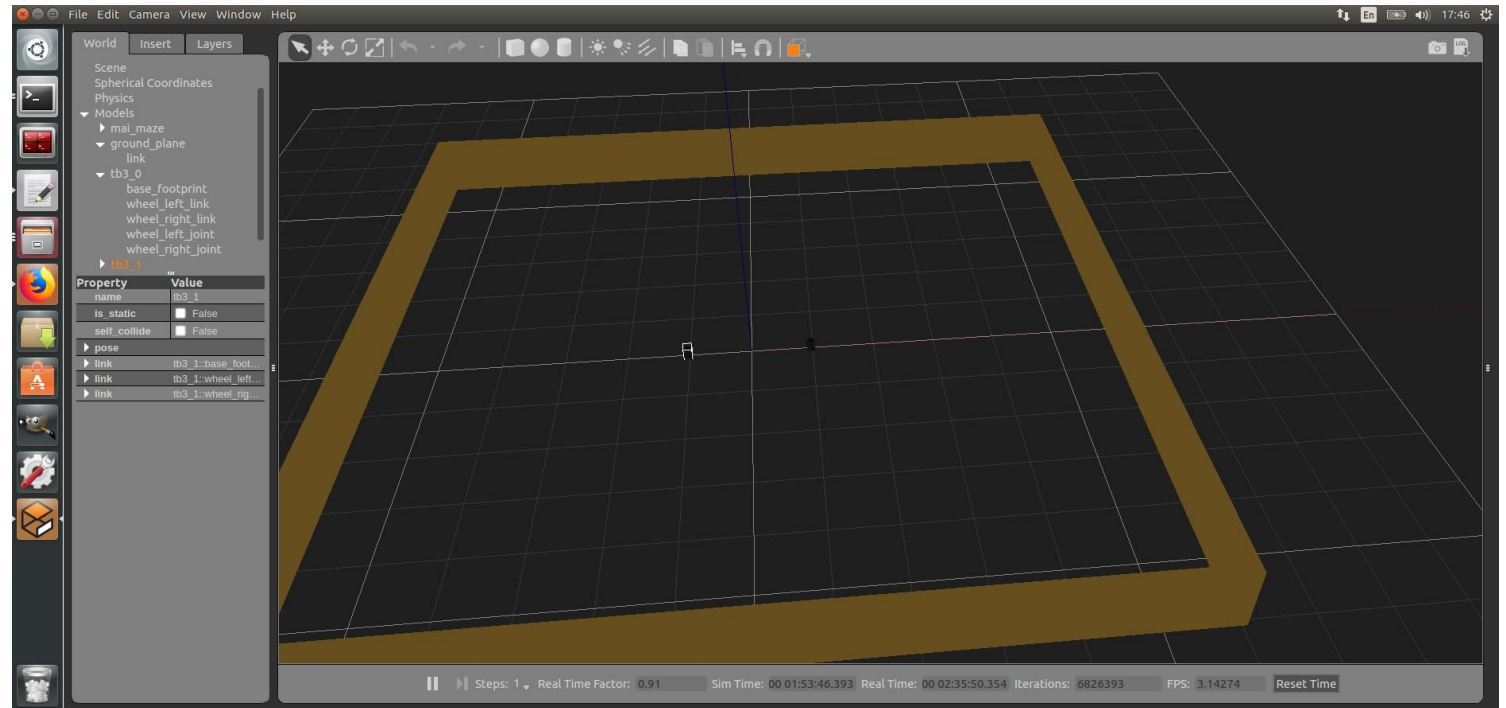
Outline

- Problem formulation
- Features
- Experiments
- Future work

Problem formulation



- outlet
- police
- thief
- pressure plate



Problem formulation

The Q function is approximated by some features.

$$Q(s, a) = \sum_i w_i f_i(s, a)$$

w:weighting vector

f:features

We use gradient descent to find w

$$w_i \leftarrow w_i + \alpha \left[R(s) + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \frac{\partial Q(s, a)}{\partial w_i}$$

Problem formulation

(Police)

Given

10*10 map outlet1(-5, 5) outlet2(5, 5)

Pressure plate: $[3*s, 3*t]$ $s, t \in (-1, 0, 1)$

robot size: $0.3*0.3$ (meter²)

State: (i, j) $-5 \leq i \leq 5, -5 \leq j \leq 6$ $i, j \in \mathbb{R}$

Action: (Up, Left, Down, Right, Stay)

Reward: 100 if police catch thief

 -100 if thief escape

 -0.1 go to the next state

Discount: 0.9

Problem formulation

(Thief)

Given

10*10 map outlet1(-5, 5) outlet2(5, 5)

Pressure plate: $[3*s, 3*t]$ $s, t \in (-1, 0, 1)$

robot size: $0.3*0.3$ (meter²)

State: (i, j) $-5 \leq i \leq 5, -5 \leq j \leq 5$ $i, j \in \mathbb{R}$

Action: (Up, Left, Down, Right)

Reward: -100 if police catch thief

100 if thief escape

-0.1 go to the next state

Discount: 0.9

Features

(Police)

feature1:constant

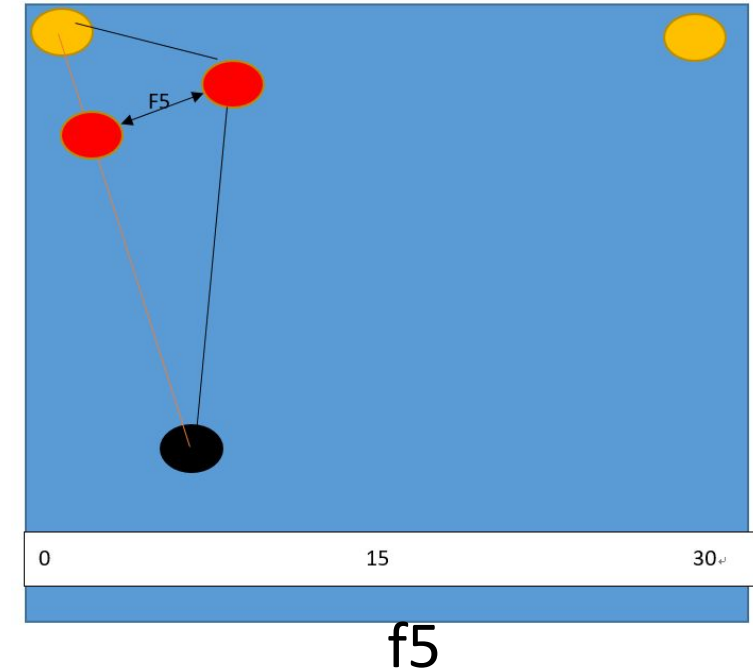
feature2:the minimum distance to the two outlets

feature3:x coordinate

feature4:y coordinate

feature5:the vertical distance from the police to the line between the thief and the exit

feature6:the distance of the police and the predict thief position



Features

(Thief)

feature1:constant

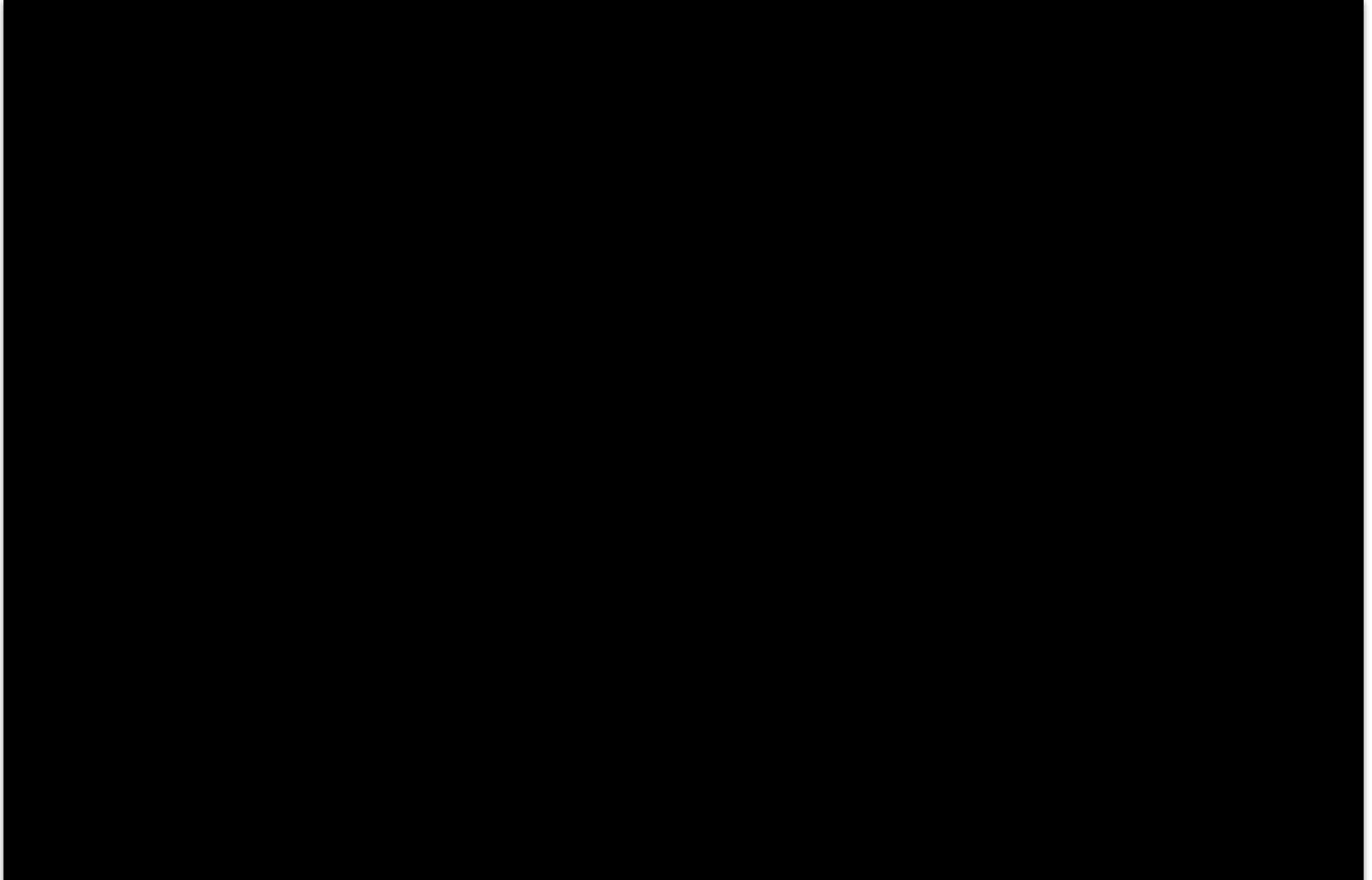
feature2:the minimum distance to the two outlets

feature3:x coordinate

feature4:y coordinate

feature5:the number of times that stay on the pressure plate

Experiments



Future

- Simulation
- Adjustment
 - feature
 - parameter
- Map reconstruct