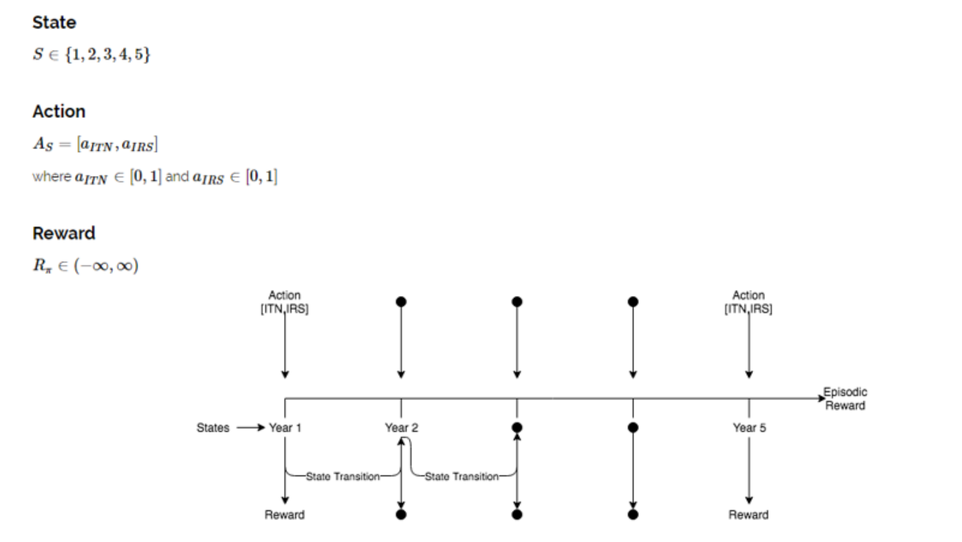
Reduce Search Space On Competition Policy Learning for Malaria Control

0. Problem Definition



State

$S \in \{1,2,3,4,5\}$

Action

$A_S = [a_{ITN}, a_{IRS}]$

where $a_{ITN} \in [0,1]$ and $a_{IRS} \in [0,1]$

Reward

$R_\pi \in (-\infty, \infty)$

- environment about this competition
  - only 20 opportunities to interact with the environment
  - Unreachable final evaluation environment when submitting final solution

1.1 Collect high score strategy - Using Q-learning

- Q-learning

State $s \in \{1,2,3,4,5\}$

Action $A_s = [a_i, 1 - a_i]$, $a_i \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$

epochs Run 1000 epochs

SARSA SARSA performence better than Q-learning

1.2 Collect high score strategy - Using GA

- Genetic Algotrithm

Initialization $a_i \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$

epochs Run 200 epochs

Mutate 1. random . 2. change x to $1 - x$

Crossover Set $A_s = [a_i, b_i]$, $T = A_1, A_2, \cdot, A_5$, let crossover point is A_3 the operation is

$$T = A_1, A_2, A_3, A_4, A_5$$
$$\Downarrow$$
$$T = A_3, A_4, A_5, A_1, A_2$$

2. Analyze high performance policy

- Collect high score strategy
  - $1 : [0.6, 0.2]$, $2 : [0.0, 1.0]$, $\cdots$, $5 : [0.6, 0.8]$
  - $1 : [0.2, 1.0]$,$2 : [1.0, 0.0]$,$\cdot\cdot\cdot$, $5 : [0.0, 0.5]$
  - $\cdots$
  - $1 : [1.0, 0.0]$, $2 : [0.0, 0.8]$, $\cdots$, $5 : [0.0, 1.0]$
- Analyze
  - $A_i \approx 1$
  - $|a_{i+1} - a_i| = 1$, $|a_{i-1} - a_i| = 1$
  - $|b_{i+1} - b_i| = 1$, $|b_{i-1} - b_i| = 1$
  - $\sum A_i = 5$
  - $a_i + b_i \approx 1$
- Test environment by hands
  - when $A_i = [1,0]$,$A_i = [1,0]$, the reward of A_{i+1} will be 0
  - when $A_i = [0,1]$,$A_i = [0,1]$, the reward of A_{i+1} will be 0
  - when $A_i = [1,0]$,$A_i = [0,1]$, the reward $A_{i+1} \approx 100$
  - when $A_i = [0,1]$,$A_i = [1,0]$, the reward $A_{i+1} \approx 100$

3.1 Reduce search space - Q-learning

- Q-learning

Initialization $a_i + b_i \approx 1$, $|a_{i+1} - a_i| \geq 0.6$, so as b_i

Action $A_s = [a_i, 1 - a_i]$, $a_i \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$, Check random policy $|a_{i+1} - a_i| \geq 0.6$, so as b_i

Random Set more possibility to choose random policy at first epochs

policy For each $A_i = [a_i, b_i]$, we can set $a_i = 0$ or $b_i = 0$ to reduce search space, So the policy can look like $[?,0]$,$[0,?]$,$\cdot\cdot\cdot$,$[?,0]$

3.2 Reduce search space - GA

- Q-learning

Initialization $a_i + b_i \approx 1$, $|a_{i+1} - a_i| \geq 0.6$, so as b_i use policy like $[?,0]$,$[0,?]$,$\cdot\cdot\cdot$,$[?,0]$

mutate For mutate opertaion, force $|a_{i+1} - a_i| \approx 1$, so as b_i , force $a_i + b_i \leq 1.4$

Huang Zi-Kuan , Xiao Jing-Jing

hzk1201@gmail.com , jingjingxiao.edu@gmail.com

Nation Cheng Kung University