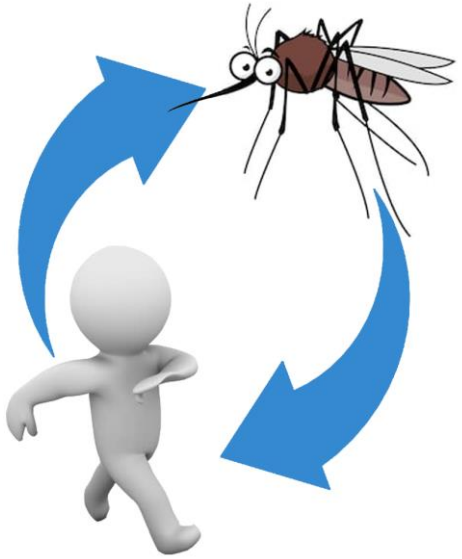
Minimizing the Spread of Drug Resistant Malaria using RL

AUGUSTE LALANDE

Motivation

- RL for social good
- Interesting problem:
  - Malaria is still very much a problem (731,000 deaths 2016)
  - Intermittent Preventive Treatment (IPT) is a malaria control strategy in which asymptomatic individuals are given a full curative dose of an antimalarial medication at specified intervals.
  - This makes a negative feedback loop decreasing the prevalence of the disease.
  - However, like over-prescription of antibiotics, this can lead to a rise in the drug resistant form of the disease.



- Interesting environment:
  - Partially observable state
  - Necessity for transfer learning

Environment

Latent State	
S	Susceptible population
I_s	Infected (sensitive strain)(symptomatic)
I_a	Infected (sensitive strain)(asymptomatic)
J_s	Infected (resistant strain)(symptomatic)
J_a	Infected (resistant strain)(asymptomatic)
T_s	Treated (infected symptomatic)
T	IPT treated (not-infected)
T_a	IPT treated (asymptomatic infected)
R	Temporarily immune

Observable State	
I_s+J_s	Infected (symptomatic)
T_s	Treated (symptomatic)
$T+T_a$	IPT Treated (asymptomatic)
$S+I_a+J_a+R$	Asymptomatic other

Action:	IPT rate (treatments per person per day)
Reward:	Number of (symptomatic) infected people at each time step

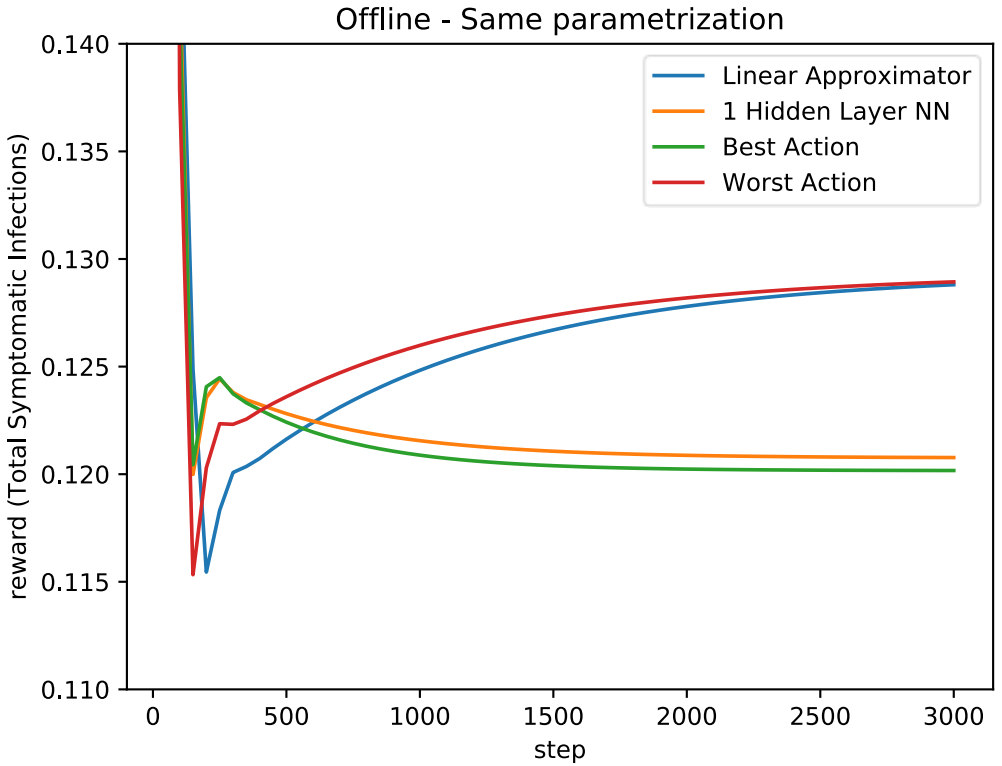
Evaluation

Two Models:

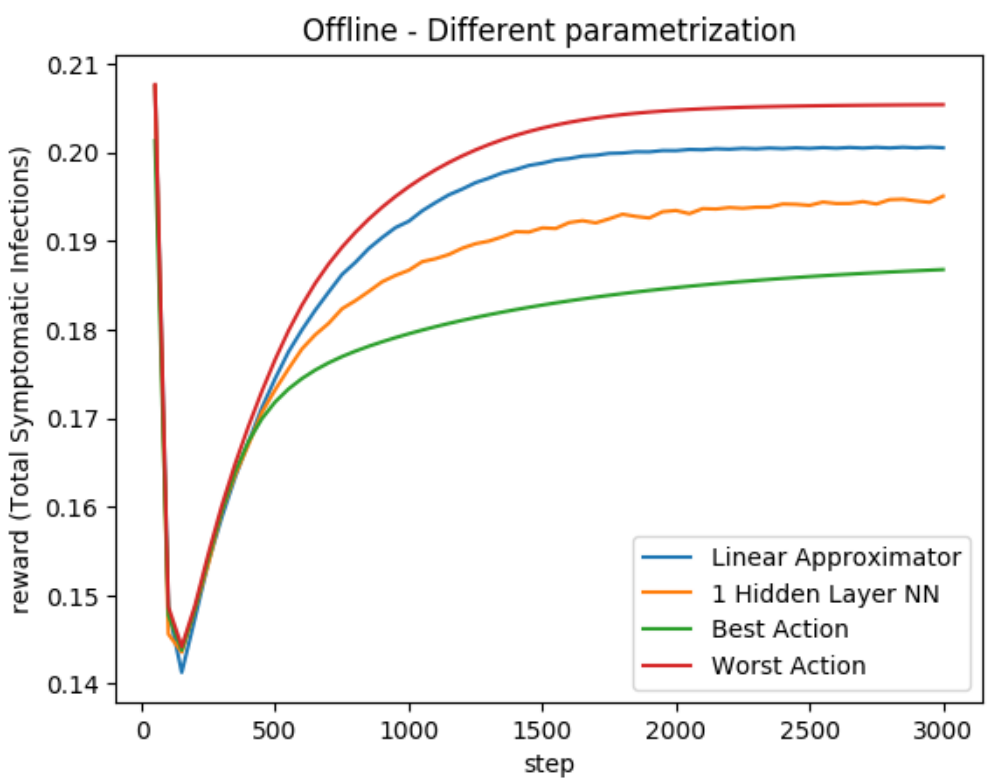
1. Linear Function Approximation
2. (Shallow) DQN with one hidden layer

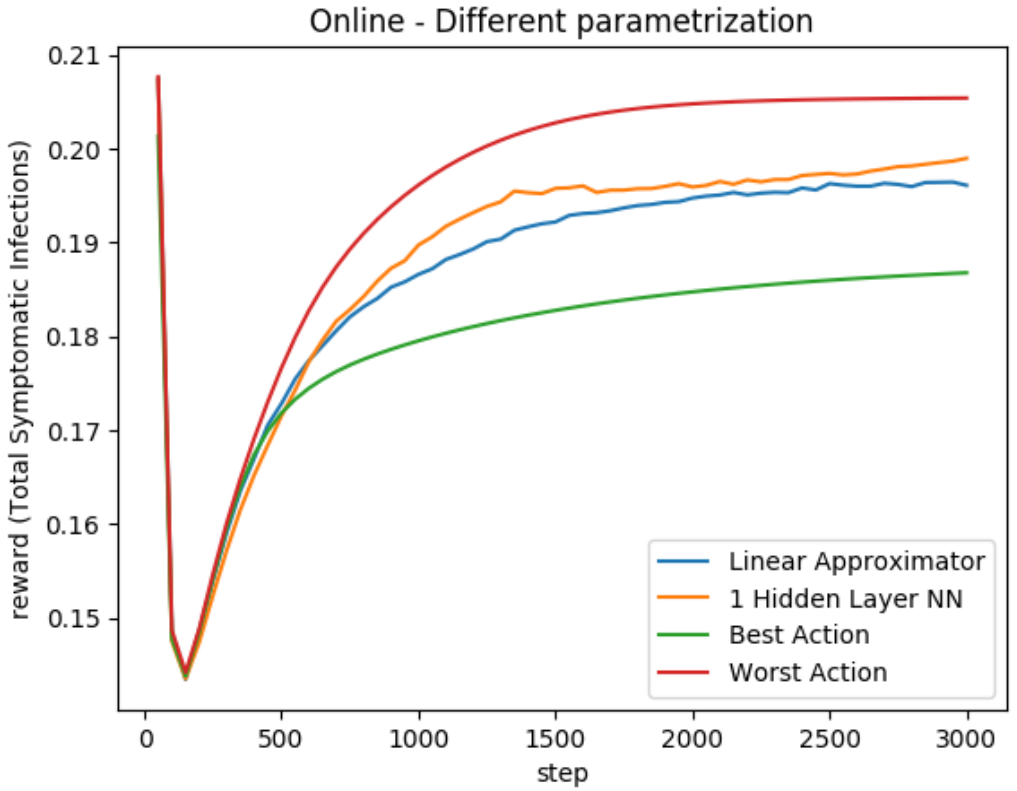
Three scenarios:

1. Offline learning with the same model parametrization
2. Offline learning with different model parametrizations
3. Online learning



Evaluation


Offline - Different parametrization
reward (Total Symptomatic Infections)
step
0.21
0.20
0.19
0.18
0.17
0.16
0.15
0.14
0
500
1000
1500
2000
2500
3000
Linear Approximator
1 Hidden Layer NN
Best Action
Worst Action


Online - Different parametrization
reward (Total Symptomatic Infections)
step
0.21
0.20
0.19
0.18
0.17
0.16
0.15
0
500
1000
1500
2000
2500
3000
Linear Approximator
1 Hidden Layer NN
Best Action
Worst Action

Future Work

- Improve environment (currently highly sensitive)
- Longer state sequence for learning
- Better hyperparameter tuning
- Better models (actor critic)