LOLs team solution for KDD Cup Reinforcement Learning 2019

Van Bach Nguyen, Bao Long Vu, Mohamed Karim Belaid, Yufeng Li

University of Passau, Germany

Problem

Malaria is dangerous disease in Africa. Human are taking 2 interventions to prevent it:

- Insecticide spraying (IRS).
- Distributing bed nets (ITN)

Our task is to find out what is the best coverage range for each intervention for each year without spending too much cost. Malaria problem in sub-Saharan Africa has been simulated by a program called OpenMalaria (https://github.com/SwissTPH/openmalaria/wiki). This OpenMalaria simulation provides an environment that returns score(reward) when we choose some actions as input.

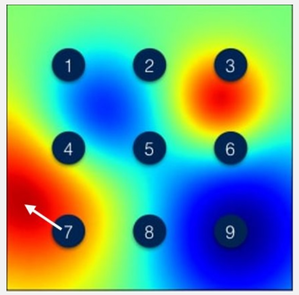
Introduction

This problems is treated as sequence decision making problems, action is a combination of only two possible interventions :Insecticide spraying (IRS) and distributing bed nets (ITN). $ITN \in [0, 1]$ and $IRS \in [0, 1]$. Action values between 0 and 1 describe a coverage range of the intervention for a simulated human population. In this problem, observations for the challenge models occurs over a 5 years time frame and each year of this time frame may be considered as the *State* of the system. With the possibility to take one Action for each State. This temporal State transition is fixed and as such not dependent on the Action taken. Our task is to find actions for series of 5 years (5 state) to maximize average rewards.

Solution

Due to limit number of evaluations (100), we decided to discretize the continuous action spaces, this means we set the resolution for the action spaces to 0.1, so, action spaces has size 100. And we also decide to break the sequence of action by finding the best action for the first year, then apply Q-learning for other years. Our algorithm is as following:

1. Using grid search and hill climbing to maximize the environment for the first year.
   - We used 16 evaluation for grid search, find the best action
   - From the best action location, we use 4 evaluations to search around that action, try to follow the direction that gives us higher reward



2. Applying Q-learning for 4 remaining years using 80 evaluation left.
   - learning rate = 1/number of visited
   - Exploration rate: $\epsilon = 0.8$, decrease over iterations

Result

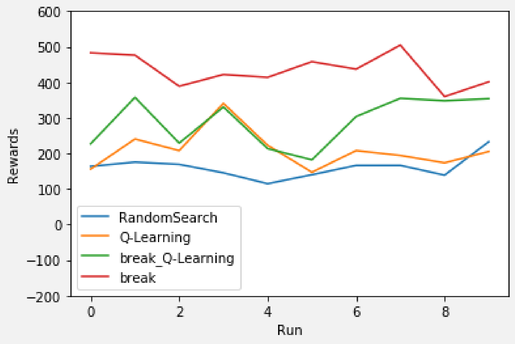


Figure: Environment 1

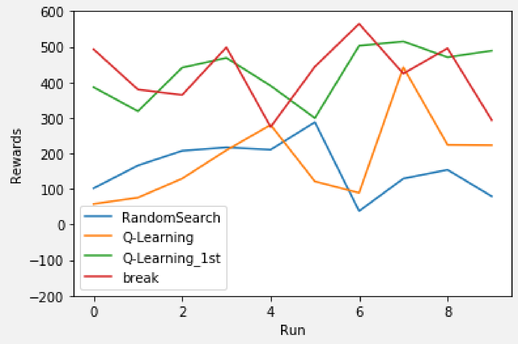


Figure: Environment 2

Conclusion

This problem is really challenge when the evaluation is very limited. Most of reinforcement learning algorithms only work when we have enough evaluations, but in this challenge, only 20 episodes are allowed.

Additional Information

We have tried other method such as Policy Gradient, Bayesian Optimisation, Genetic Algorithm, but if we do not have any special solution like breaking the sequence, maximizing the first year rewards, the results will be very bad.

References

[1] RS Sutton and AG Barto.
Reinforcement Learning.
2nd edition, 2018.

[2] Stephen Roberts Aisha Walcott-Bryant Oliver Bent, Sekou L. Remy.
Novel exploration techniques (nets) for malaria policy interventions.

Contact Information

- Email: nvbach92@gmail.com