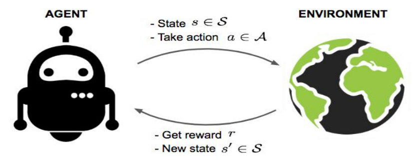
Epidemic Control With Reinforcement Learning

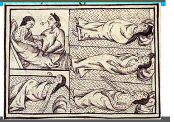
Faisal Al-Kwiliy, Ashvin Uthayakumar, Qiyao Wei, Team 2020297, Session 3

Supervisor Andreas Moshovos, Admin Nick Burgwin



Background and Motivation

[1]



- Epidemics date back to 1704 with smallpox
- A lot of research has been dedicated to understanding and solving SIR models
- Past epidemic data don't convey the full details of the epidemic

Project Goal

The problem is there is a lack of research in utilizing Reinforcement Learning to achieve epidemic control. In order to provide a systematic Reinforcement Learning solution to this problem, we aimed to achieve the control of epidemics using reinforcement learning while matching or surpassing GEKKO by achieving similar or lower speed and penalty. [2]

Key Requirements

9 Requirements:

1. Capable of running on a Macbook
2. Must be solved using RL
3. Must not deviate more than 10% for each data point compared to the epidemic models that we are using as a reference
4. Solutions obtained via setting all actions to zero and without the use of RL must be within 10% deviation for each data point
5. The weighted sum of population death penalty and control measure penalty must be lower in value than heuristics such as do-nothing (setting all actions to zero)
6. Match or surpass existent techniques such as optimal control with GEKKO based on having a lower value for the weighted sum of penalties
7. Must produce a solution as fast as GEKKO in wall clock time
8. ODEs with control parameters should behave the same as the original SIR ODEs when the control parameters are all set to zero
9. Setting control parameters to positive values should always limit the negative effects of the epidemic

Challenges

Interfacing RL with SIR

- Lots of moving pieces in the Gym environment
- Very time-consuming to implement

Quantifying the cost of actions and control variables

- Output isn't always sensible
- Many parameters in need of tuning

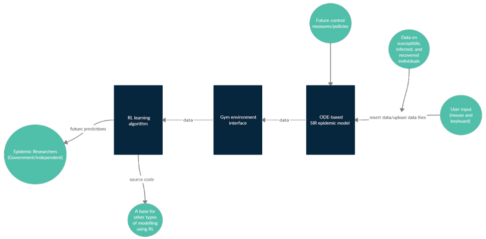
Limited number of samples/usable data

- Difficult to know if we overfitted our data
- Need more data to ensure greater accuracy

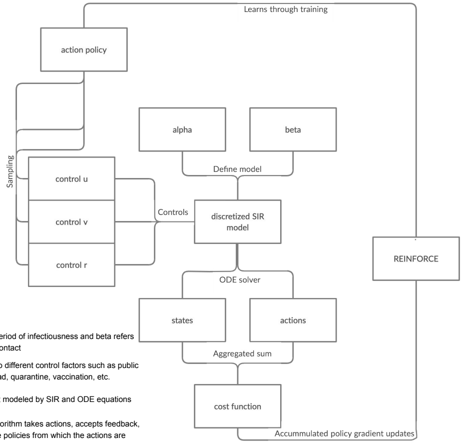
[1]-http://1704.deerfield.history.museum/popups/background.do?shortName=expEpidemics

[2]-https://memoori.com/will-mass-surveillance-become-a-leading-epidemic-control-technology/

System Level Overview

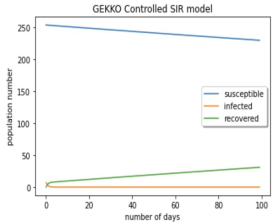


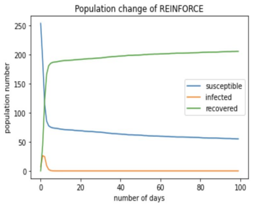
System Block Diagram



- Alpha refers to period of infectiousness and beta refers to frequency of contact
- u, v, and r refer to different control factors such as public information spread, quarantine, vaccination, etc.
- Use environment modeled by SIR and ODE equations
- REINFORCE algorithm takes actions, accepts feedback, and improves the policies from which the actions are decided
- Trains algorithms by taking control measures, applying them on the environment, and learning from environment feedback

Experiment Results





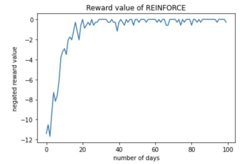

```
final reward value is 0.0003
```

- GEKKO is the benchmark algorithm, and REINFORCE is our implementation
- Both algorithms produce reasonable solutions, but GEKKO has a better reward value (see appendix)
- This is because GEKKO figures out to immediately control the infected population, whereas REINFORCE did not learn that

Conclusion

- We have devised a fully-functional RL algorithm for real-world epidemic control
- Even though our algorithm did not outperform GEKKO in reward value, it was faster than GEKKO
- We are highly optimistic about more sophisticated RL algorithms outperforming GEKKO entirely
- Moreover, given more data and training time, even our simplistic REINFORCE algorithm might defeat GEKKO because of its data-driven nature

Appendix: Reward Plot




```
final reward value is 0.2920616373302047
```