Reinforcement Learning Seminar

Boyeon Kim

Department of Mathematics, School of Mathematics and Computing
Mathematics
Yonsei University

May 12, 2023

SEIAR Optimal control

- Previous experiments have shown that for one control,
  we can find optimal control using reinforcement learning.
- By introducing a new problem that is one control,
  we will try to do optimal control in the following two ways.
- Goal1 : DQN
- Goal2 : PPO

Mathematical models

- The influenza model : SEIAR model
- J.Kim et.al.,*Constrained optimal control applied to vaccination for influenza*,2016

$$
\begin{aligned}
S'(t) &= -\beta S(t)\Lambda(t) - \psi\nu(t)S(t) \\
E'(t) &= \beta S(t)\Lambda(t) - \kappa E(t) \\
I'(t) &= p\kappa E(t) - \alpha I(t) - \tau I(t) \\
A'(t) &= (1-p)\kappa E(t) - \eta A(t) \\
R'(t) &= f\alpha I(t) + \tau I(t) + \eta A(t) + \psi\nu(t)S(t)
\end{aligned}
$$

with $\Lambda(t) = \epsilon E(t) + (1-q)I(t) + \delta A(t)$

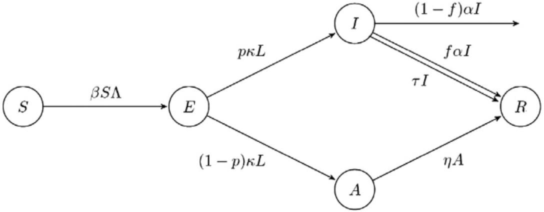

Fig. 1. Flow chart for the SEIAR model.

SEIAR model parameters

Parameter	Description	value
ϵ	Infectivity reduction factor for the exposed	0
q	Contact reduction by isolation	0.5
δ	Infectivity reduction factor for the asymptomatic	0.5
p	Fraction of developing symptoms	0.667
κ	Transition rate for the exposed	0.7143 / day
f	Complement to fatality rate (1 - fatality rate)	0.999
α	Recovery rate for the (symptomatic) infective	0.1667 /day
η	Recovery rate for the asymptomatic	0.1667 / day
τ	Antiviral treatment rate	0 / day
ψ	Efficacy of vaccination	70%
β	Transmission rate	6.3346e-08
R_0	Basic Reproduction number	1.9

- $S0 = 5e07, \quad E0 = 0, \quad I0 = 1, \quad A0 = 0, \quad R0 = 0$

Goal in paper

- The goal is to **minimize the number of people who become infected** at a **minimal efforts of vaccination**.
- The objective functional is given by

$$J(\nu) = \int_0^T PI(t) + Q\nu^2(t)dt$$

with inequlity constraints $0 \leq \nu(t) \leq 1, \quad \nu(t)S(t) \leq \nu_{max}, \quad z(t) = \int_0^T \nu(t)S(t) \leq \nu_{total}$

- ν_{max} : The maximum daily vaccination
- ν_{total} : Vaccine coverage
- $z'(t) = \nu(t)S(t) \quad z(0) = 0 \quad z(T) \leq \nu_{total} \quad z'(t) \leq \nu_{max}$

penalty method

- The constrained optimal control → into unconstrained optimization problems.
- The objective penalty function

$$
\begin{aligned}
J_p(\nu) = \int_0^T & PI(t) + Q\nu^2(t) \\
& + \mu_1(\nu(t)S(t) - \nu_{max})^2 H_1(\nu(t)S(t) - \nu_{max}) \\
& + \mu_2(z(t) - \nu_{total})^2 H_2(z(t) - \nu_{total})dt
\end{aligned}
$$

with Heaviside step function

$$
H_1(\nu(t)S(t) - \nu_{max}) = \begin{cases} 0 & if \quad \nu(t)S(t) \leq \nu_{max} \\ 1 & if \quad \nu(t)S(t) > \nu_{max} \end{cases}
$$

$$
H_2(z(t) - \nu_{total}) = \begin{cases} 0 & if \quad z(t) \leq \nu_{total} \\ 1 & if \quad z(t) > \nu_{total} \end{cases}
$$

with $z(t) = \int_0^T \nu(t)S(t)dt$

- $z'(t) = \nu(t)S(t) \quad z(0) = 0 \quad 0 \leq \nu(t) \leq 1$

Reinforcement Learning

Set the environment for DQN

- $\nu_{max} = 5e05(0.01 * S0)$
- $\nu_{total} = 5e06(0.1 * S0)$
- Observation space : 5 (S, E, I, A, R)
- Action space : 2 (0 or 1)
- Action
  - ν : the number of vaccinated people
  - action $= 0 \rightarrow \nu = 0$
  - action $= 1 \rightarrow \nu = \nu_{max}$
  - $\nu(\text{ratio}) = \nu/\nu_{max}$
- Reward design
  - case1)
    - \- I - ν
    - if sum(ν) $> \nu_{total}$, penalty reward: - 10000
  - case2) Similar to the paper
    - -PI - Q$\nu(ratio)^2$
    - if $\nu(ratio) * S(t) > \nu_{max}$, penalty reward: -10000
    - if sum(ν) $> \nu_{total}$, penalty reward: -10000
    - P, Q : proper weight

Gaol1 : DQN

SELAR optimal: DQN

- Error

```
⊗ (epi) boyeon@gimboyeon-ui-MacBookAir Boyeon % /opt/anaconda3/envs/epi/bin/python /Users/boyeon/research/ezlab-rl/seiar_ppo.py
   10%                                                   49,935/500,000  [ 0:00:31 < 0:04:39 , 1,615 it/s ]
  zsh: segmentation fault  /opt/anaconda3/envs/epi/bin/python
  /opt/anaconda3/envs/epi/lib/python3.10/multiprocessing/resource_tracker.py:224: UserWarning: resource_tracker: There appear to be 1 leaked semaphore objects to clean up at shutdown
    warnings.warn('resource_tracker: There appear to be %d '
```

Next

- Google colab 연동(환경 맞추기)
- 500,000step까지 작동 확인
- Reward Design : Case1, Case2 Check!