Reinforcement Learning for Finding COVID-19 Policy with Various Actions

2021. 1. 29. 중간 보고

박혜린

개요

갑작스럽게 시작된 COVID-19 pandemic 상황은 벌써 1년 간 지속되었고, 이에 따라 최적의 방역 정책을 찾으려는 노력은 각국에서 계속되고 있다. 최적의 정책을 학습한다는 점에서 강화학습을 이용해 최적의 방역 대책을 찾는 연구가 꽤 진행되었다. 하지만 전염병 확산 연구의 경우, 실제 모든 영향 요인을 고려하기 어렵다는 점에서 환경 구축 부분이 쉽지 않다. 이에 연구 별 환경 설정 부분이 천차만별이다. 따라서 우리나라의 방역 정책을 반영한 COVID-19 강화학습 모델 구현을 이번 연구의 목적으로 한다. 그 과정에서 COVID-19라는 실제 문제에 적용된 알고리즘 간의 결과 차이를 비교해보고, 그 코드를 github에 공유하여 결과물을 공유한다.

문제 정의

<연구 주제>

Reinforcement Learning

for Finding COVID-19 Policy

with Various Actions

1) 최적의 정책을 학습하는 강화학습을 이용해,
우리나라의 데이터를 바탕으로 COVID-19에 적용시켜본다.

2) 다른 국가들과 달리 단순 봉쇄(lockdown) 대신,
세밀한 단계별 방역 대책을 시행한 우리나라에 맞춰
agent의 action을 다양화하여 구현해본다.

3) 다양한 강화학습의 알고리즘들에 대해 알고, COVID-19
문제에 맞춘 각 알고리즘 실행 및 그 결과를 비교해본다.

기존 방법

* 코로나 강화학습과 관련된 연구는 대부분 이 논문과 비슷함

논문	*EpidemiOptim*: A Toolbox for the Optimization of Control Policies in Epidemiological Models, 2020.10.9.	Optimal Policy Learning for COVID-19 Prevention Using Reinforcement Learning, 2020.9.16.
Data	프랑스 (인구수, GDP 값 계산)	바탕 데이터 없음. 그저 reward 값으로 알고리즘 별 결과 비교.
Environment – Epidemiological model	SEIRAH model (SEIR 모델에 무증상자(A), 병원 치료자(H) 추가)	없음.
Environment – reward	Health (사망자 수) & economy (GDP loss) cost 최소화	임의로 지은 reward 함수 값 최대화
Agent – Action	Lockdown(봉쇄) on/off – 2가지	Testing(검사 수), Sanitization, Lockdown 비율 조정
Agent – Learning algorithm	DQN, NSGA-ii (경제학의 최적화 방법)	Q러닝, SARSA, DQN, DDPG 비교

장단점

+ 실제 데이터 적용
+ 전염병 확산 모델 고려
- 2가지 뿐인 action 수
- 강화학습의 DQN만 적용
+ 코드 공개

VS.

- data 사용하지 않아 실제문제에 적용하기 아직 어렵
- 전염병 특성 고려 X
+ 다양한 코로나 정책 고려
+ 여러 알고리즘 구현/비교

진행 방법

기존 연구들

- '강화학습' 자체에만 집중해 전염병 확산 특성에 대한 고려가 부족하여 아직까지는 실제 문제 상황에 적용시키기는 어렵거나,
- 전염병 확산 특성이 너무 복잡한 탓에 간단한 action을 하는 강화학습 적용에 그침

GOAL 공개된 코드(EpidemiOptim)를 이용하여 우리나라 상황에 맞는 강화학습 모델 구현

1. Data : 프랑스 → 우리나라
2. Action : lockdown on/off → 단계별 사회적 거리두기 (action 다양화)
3. Epidemiological model : SEIRAH → SQEIR (앞선 2의 action의 변화 따라 필요한 수정)
4. Learning algorithm : DQN → DDPG (DDPG가 DQN보다 continuous space에서 더 효과적인 것으로 알려져 있기 때문)

   ⇒ DQN과 DDPG 결과 비교

* (if possible) Prediction에 좋은 확률적 프로그래밍과의 결합도 생각해보았으나, 무리일 것으로 판단됨

현재 진행 상황

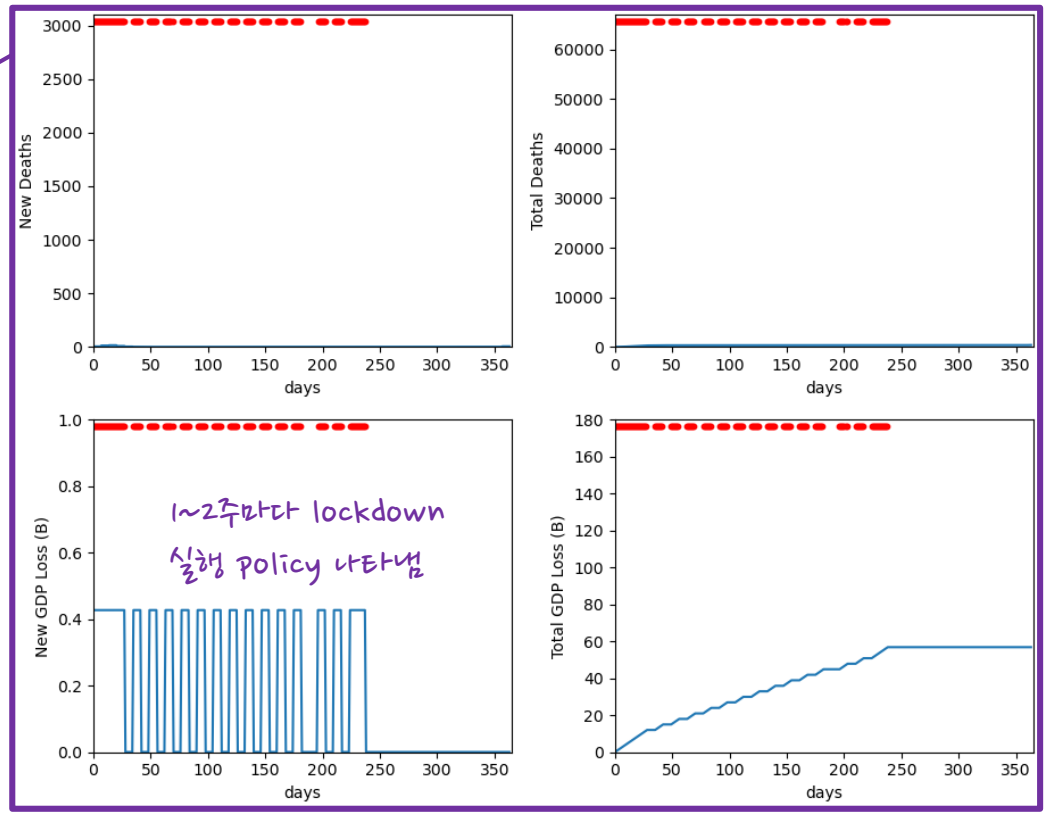
- 논문 조사: 앞선 2개의 논문이 가장 바탕이 되기에 구현 방식 및 알고리즘 살펴봄

- 기본 코드의 바탕이 될 EpidemiOptim(github) 분석
  - 파일/코드 구조 파악
  - 총 4개의 알고리즘 (DQN, DQN 변형 2개, NSGA-ii(경제학의 알고리즘)) 코드 및 결과 확인

  * 필요한 Pytorch 공부

- Data 수집
  - coronaboard.kr : kr_daily.csv(우리나라 코로나 현황)
  - KT빅데이터플랫폼: fpopl.csv(행정동별 유동인구 데이터) 등

Example - DQN



예상 결과물

✓ 우리나라의 상황에 맞는 코로나 정책 관련 강화학습 환경 구현

- 확진자/사망자 수 등의 규모를 우리나라에 맞게 알 수 있음

- 우리나라의 방역 지침 기준에 따른 정책 학습 기대

* 구현 완료한 코드(in Python)는 github에 공유할 예정

✓ COVID-19라는 실제 문제 상황에서의 알고리즘 간 결과 비교

- DQN과 DDPG의 결과 비교 예정