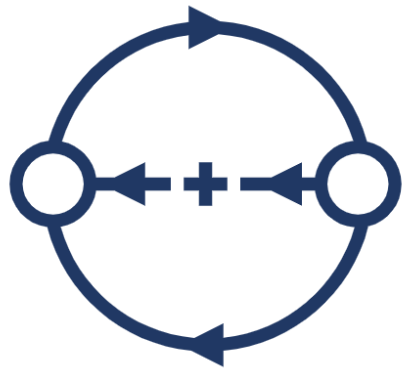


Reinforcement Learning for Finding COVID-19 Policy with Various Actions

2021. 2. 23. 최종 발표

동계 연구연수생 박혜린

Introduction



Reinforcement Learning

: Learn optimized policy
itself
(ex) 주가 예측



BUT

전염병 확산 연구

- 실제 영향 요인을 모두 고려하기 어렵다!
- 그래서 환경 구축이 쉽지 않음
- 그 결과 연구 별 환경 설정 부분이 천차만별

Introduction

<연구 주제>

Reinforcement Learning
for **Finding COVID-19 Policy**
with **Various Actions**

1) 최적의 정책을 학습하는 **강화학습**을 이용해,
우리나라의 데이터를 바탕으로 **COVID-19에 적용**시켜본다.

2) 다른 국가들과 달리 단순 봉쇄(lockdown) 대신,
세밀한 단계별 방역 대책을 시행한 **우리나라**에 맞춰
agent의 **action**을 **다양화**하여 구현해본다.

3) COVID-19 문제에 맞춰 모델링한 **강화학습 실행** 및 그
결과를 **분석**해본다. 또한 구현 코드를 github에 공유한다.

Problem

* 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

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

장단점
VS.

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

Problem

* 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.

기존 연구들

- '강화학습' 자체에만 집중해 전염병 확산 특성에 대한 고려가 부족하거나,
 - 전염병 확산 특성이 너무 복잡한 탓에 간단한 action을 하는 강화학습 적용에 그침
- ➡ 아직까지는 복잡한 실제 문제 상황에 적용시키기 어려움



GOAL

공개된 코드(**EpidemiOptim**)를 이용하여,
우리나라 상황에 맞는 COVID-19 강화학습 모델 구현

Key Idea

GOAL 공개된 코드([EpidemiOptim](#))를 이용하여, 우리나라 상황에 맞는 COVID-19 강화학습 모델 구현

1 Data: 프랑스 → 우리나라

* 출처: 행정안전부, 질병관리청 감염병포털, coronaboard.kr 등

2 Action: lockdown on/off (2가지) → 단계별 사회적 거리두기 (6가지)

```
actions = np.random.choice([0, 1, 2, 3, 4, 5], size=53)
```

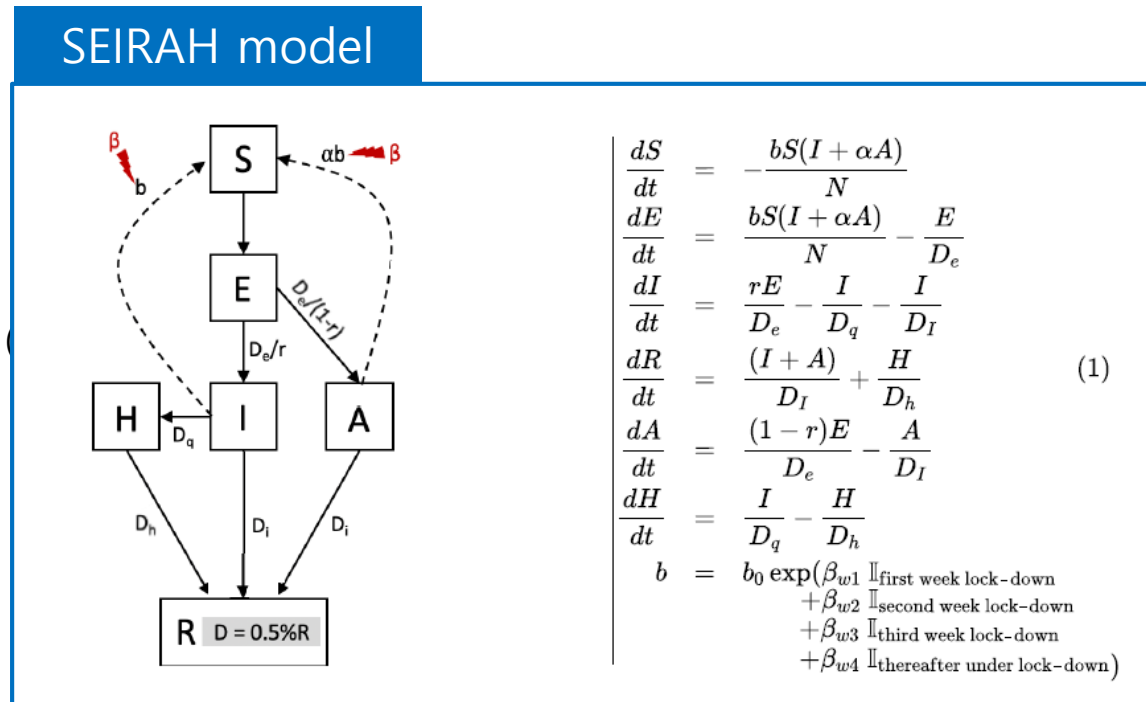
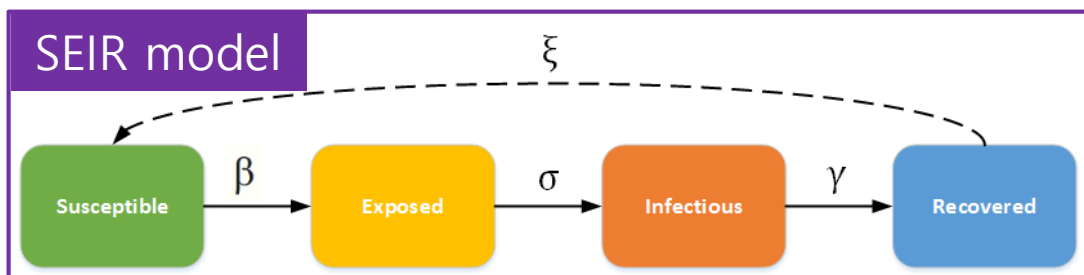
3 Epidemiological Model: SEIRAH → SQEIR

(action 의 변화에 따라 필요한 수정)

Key Idea

GOAL 공개된 코드([EpidemiOptim](#))를 이용하여, 우리나라 상황에 맞는 COVID-19 강화학습 모델 구현

- 1 **Data:** 프랑스 → 우리나라
- 2 **Action:** lockdown on/off (2가지) → 단계별
- 3 **Epidemiological Model:** SEIRAH → **SQEIR**



Key Idea

GOAL 공개된 코드([EpidemiOptim](#))를 이용하여, 우리나라 상황에 맞는 COVID-19 강화학습 모델 구현

1 **Data:** 프랑스 → 우리나라

* 출처: 행정안전부, 질병관리청 감염병포털, coronaboard.kr 등

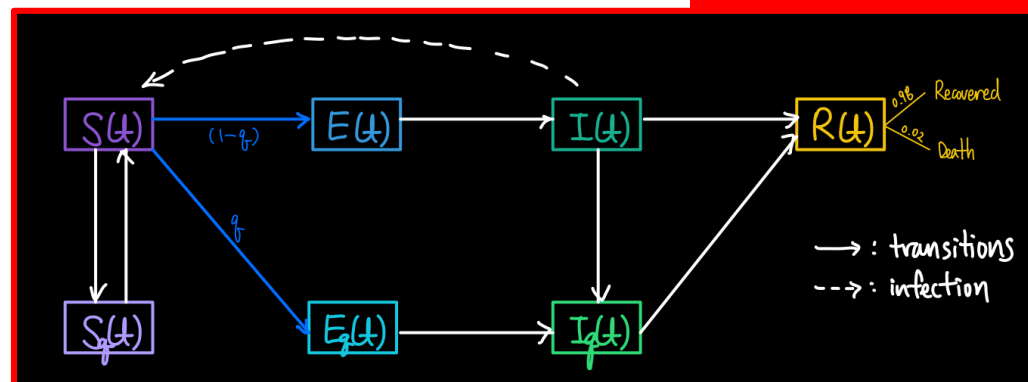
2 **Action:** lockdown on/off (2가지) → 단계별 사회적 거리두기 (6가지)

```
actions = np.random.choice([0, 1, 2, 3, 4, 5], size=53)
```

3 **Epidemiological Model:** SEIRAH → SQEIR

(action 의 변화에 따라 필요한 수정)

SQEIR model



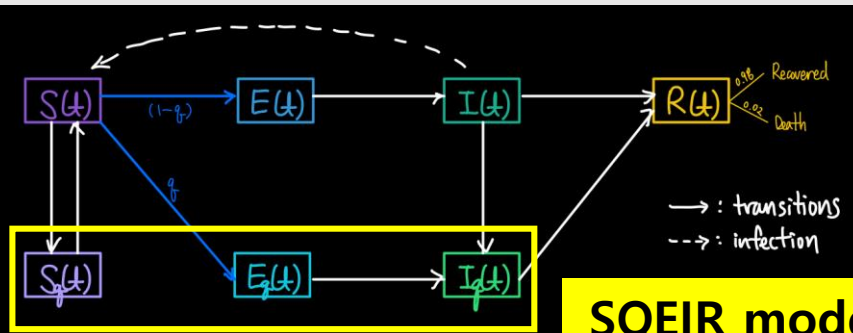
Contribution

✓ New Modeling for COVID-19

✓ RL Training for COVID-19 Situation in Korea & Analysis

Contribution

✓ New Modeling for COVID-19



SQUEIR model

$$\begin{aligned}\frac{dS}{dt} &= \left(-\frac{\beta C_0 S I}{N}\right) - \frac{(1-\beta) C_0 q \cdot S \cdot I}{N} + \frac{S_q}{14} \quad \dots \text{자격격리 기간} \\ \frac{dS_q}{dt} &= -\frac{S_q}{14} + \frac{(1-\beta) C_0 q \cdot S \cdot I}{N} \\ \frac{dE}{dt} &= \frac{(1-q) \beta C_0 S I}{N} - \frac{E}{D_e} \quad \dots \text{i.e. 잠복} \\ \frac{dE_q}{dt} &= \frac{q \beta C_0 S I}{N} - \frac{E_q}{D_e} \\ \frac{dI}{dt} &= \frac{E}{D_e} - \frac{I}{D_I} - \frac{I}{D_R} \\ \frac{dI_q}{dt} &= \frac{E_q}{D_e} + \frac{I}{D_I} - \frac{I_q}{D_I} \\ \frac{dR}{dt} &= \frac{I}{D_R} + \frac{I_q}{D_R}\end{aligned}$$

$S(t)$: 일반 사람
 $S_q(t)$: " 中 격리된 사람 (접촉자로 분류. but 감염 X)
 $E(t)$: 접촉자로 감염됨. but 격리되지 X
 $E_q(t)$: " . & 격리중
 $I(t)$: " → 증상 발현 but 아직 격리되지 X
 $I_q(t)$: 격리된 증상 있는 사람 (집 아 병원)
 $R(t)$: 테이싱 바이러스 X. (양자자 + 사망자)

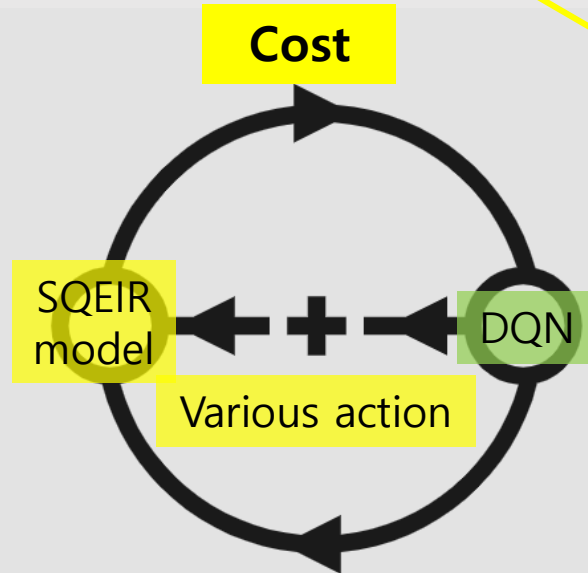
Various actions

Parameters of the SQUEIR model

Parameter	Interpretation	Value
β	Transmission rate of the virus	0.027 (2020.3.20. 기준)
C_0	Average # of contact per day	(40, 25, 20, 15, 10, 5) // 자격격리 단계
q	Quarantine rate of the contacted (ie. incubation period)	0.8
D_e	# of days to be symptomatic after be exposed	5.2
D_I	# of days to be recovered after be symptomatic	20.1
D_T	# of days need for testing then be quarantined	2.5
N	Total population	51,844,627
S_0	Initial # of susceptible	$N - (\text{아래의 나머지 2가지})$
S_{q0}	" quarantined susceptible	13261
E_0	" exposed	2000
E_{q0}	" quarantined exposed	367
I_0	" ascertained (symptomatic)	2850
I_{q0}	" quarantined symptomatic	5684
R_0	" virus removed (recovered + death)	3277 * 2020.3.23. 04

Contribution

✓ New Modeling for COVID-19



$$\bar{C} = (1 - \beta) C_{health} + \beta C_{economy}$$

$$\begin{cases} C_{health} = \text{number of death toll} \\ C_{economy} = \text{number of quarantines} \\ \beta = \text{input parameter} \end{cases}$$

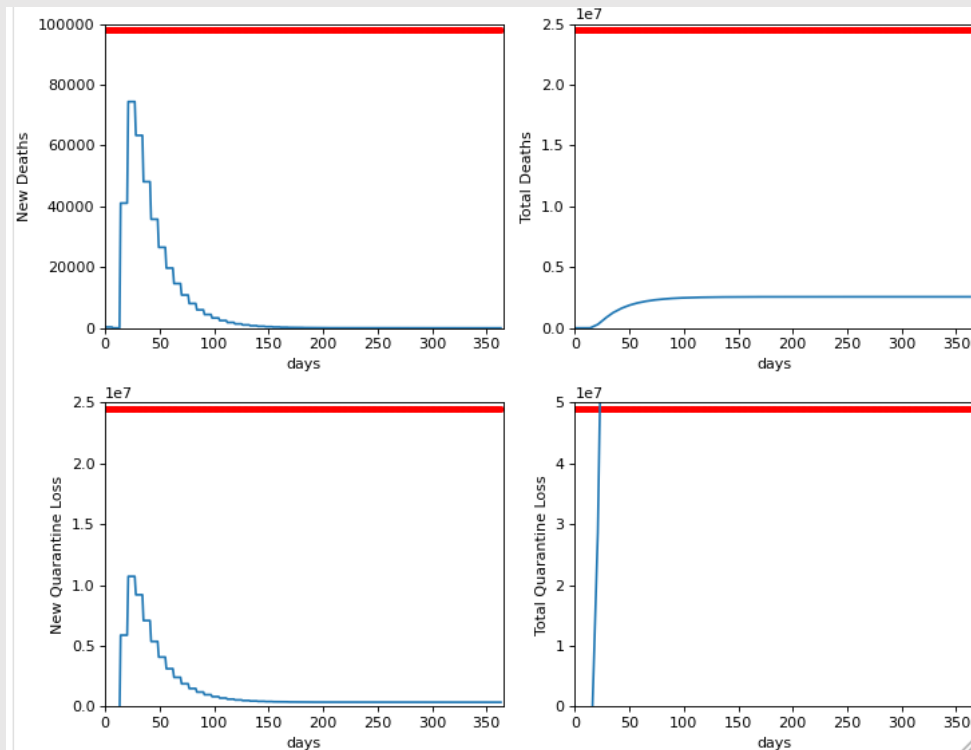
- Environment에 저장된 epidemiological model의 compartment 값(인구수) 이용 → 빠르고 간단히!
- 양면성의 cost 고려 ⇒ Learn balance!

Contribution

✓ New Modeling for COVID-19

✓ RL Training for COVID-19 Situation in Korea & Analysis

$$\bar{C} = (1 - \beta) C_{health} + \beta C_{economy}$$



1) $\beta = 0.5$

beta 0.50

Load a new DQN model for beta = 0.5

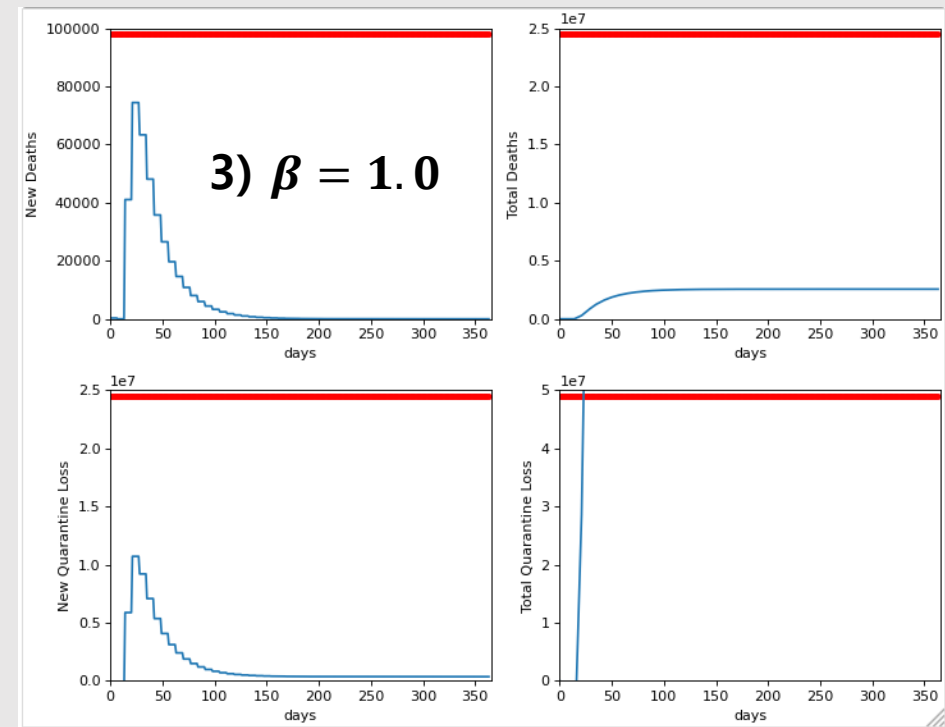
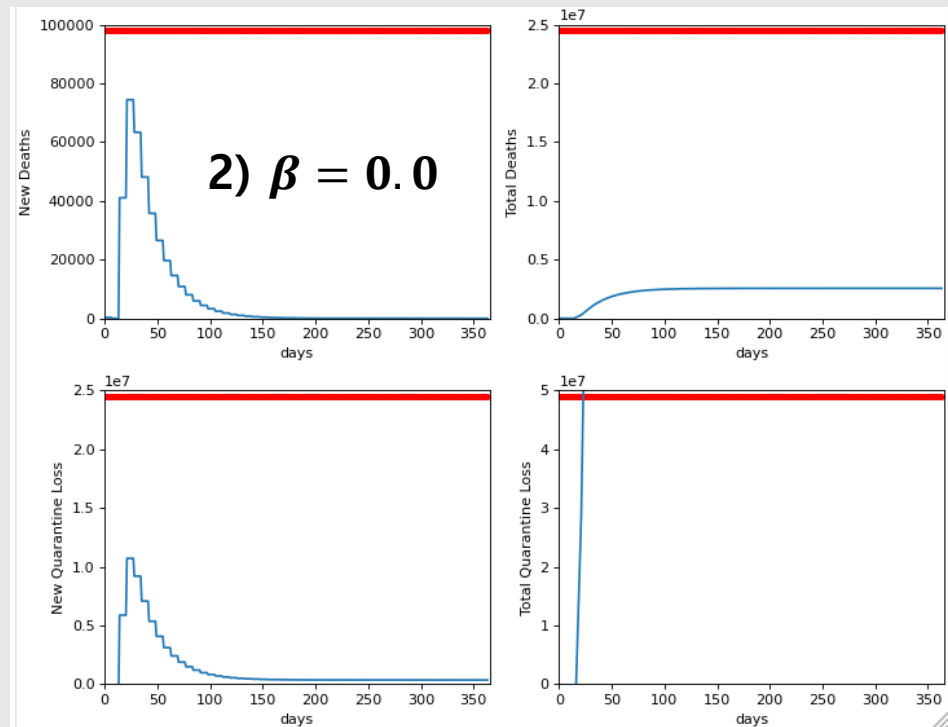
```
-----  
Eval, g: [0.5]: mean_C0: 2560665.87  
Eval, g: [0.5]: std_C0: 0.00  
Eval, g: [0.5]: mean_C1: 471476341.46  
Eval, g: [0.5]: std_C1: 0.00  
Eval, g: [0.5]: mean_agg: 364643.85  
Eval, g: [0.5]: std_agg: 0.00  
Eval score: 364643.85
```

Contribution

✓ New Modeling for COVID-19

✓ RL Training for COVID-19 Situation in Korea & Analysis

$$\bar{C} = (1 - \beta) C_{health} + \beta C_{economy}$$




Contribution

✓ New Modeling for COVID-19

✓ RL Training for COVID-19 Situation in Korea & Analysis

$$\bar{C} = (1 - \beta) C_{health} + \beta C_{economy}$$


2) $\beta = 0.0$

beta  0.00

Load a new DQN model for beta = 0.0

Eval, g: [0.]: mean_C0: 2469912.89
Eval, g: [0.]: std_C0: 0.00
Eval, g: [0.]: mean_C1: 594093249.30
Eval, g: [0.]: std_C1: 0.00
Eval, g: [0.]: mean_agg: 3799.87
Eval, g: [0.]: std_agg: 0.00
Eval score: 3799.87


1) $\beta = 0.5$

beta  0.50

Load a new DQN model for beta = 0.5

Eval, g: [0.5]: mean_C0: 2560665.87
Eval, g: [0.5]: std_C0: 0.00
Eval, g: [0.5]: mean_C1: 471476341.46
Eval, g: [0.5]: std_C1: 0.00
Eval, g: [0.5]: mean_agg: 364643.85
Eval, g: [0.5]: std_agg: 0.00
Eval score: 364643.85

3) $\beta = 1.0$

beta  1.00

Load a new DQN model for beta = 1.0

Eval, g: [1.]: mean_C0: 2562242.40
Eval, g: [1.]: std_C0: 0.00
Eval, g: [1.]: mean_C1: 444184159.36
Eval, g: [1.]: std_C1: 0.00
Eval, g: [1.]: mean_agg: 683360.25
Eval, g: [1.]: std_agg: 0.00
Eval score: 683360.25



Contribution

✓ New Modeling for COVID-19

✓ RL Training for COVID-19 Situation in Korea & Analysis

⇒ Completed new COVID-19 modeling for RL !

[code] https://github.com/hyerinshelly/RL_COVID-19_Korea

* 연구연수생을 마치며

- 강화학습에 대한 지식이 거의 없었는데, 강화학습에 대한 기본 개념부터 최근 알고리즘까지 많은 것을 김진우 연수생과 함께 탐구할 수 있어서 좋았습니다.
- 강화학습을 적용시킬 수 있는 문제에 대해 직접 생각해보고 찾아내는 과정이 처음에는 막막했지만, 연구 과정에서 가장 중요한 과정을 체험해본 것 같아 의미있었습니다.
- 실제 문제 상황의 모델링 구현을 처음 해보았습니다. 모델링의 가장 큰 틀 구성부터 파라미터들의 설정까지 직접 해보고, 3일 동안 디버깅을 해본 경험을 잊지 못할 것 같습니다.
- 비록 많은 분들의 얼굴을 많이 뵈지 못했지만, 걱정을 안고 시작한 저의 연구연수생 생활을 잘 마무리 지을 수 있도록 제 곁에서 도와주신 모든 분들께 **감사드립니다.**

감사합니다 😊

ETRI 자율형IoT연구실

2020 통계연구연수생 박혜린