

김종윤

2023-28318

협동과정 인공지능 전공

소셜컴퓨팅

과제 6 - 7~8주차

1. Network Book Exercise 16.8 - Q1, 4, 5

1.1. Q1

1.1.1. a)

The action decisions will cascade through the network. The action decision of individual 1 will be unchanged as the person do not have any information. The individual 2 will also not change the decision as the person only can be affected by individual 1 even on the previous setup.

1.1.2. b)

The individual 3 will only observe action of individual 2. The action of individual 2 is already affected by the individual 1. The decision made by individual 3 is affected by not only individual 2 but also 1. Therefore, the individual 3 can infer some observation about the individual 1 via the observation of individual 2.

1.1.3. c)

The individual 3 can view some decision made by individual 1 via individual 2's action. For instance, assume the people to deduce which the color of the ball is majority where red and blue exists. If individual 2 observed that individual 1 selected red as majority and individual 2 also picked red, then individual 3 can guess that both individual 1 and 2 picked red so the majority is red. Otherwise, individual 2 observed that individual 1 selected red as majority and individual 2 picked blue, then individual has guessed which to be the majority with own decision process that includes both individual 1's observation and 2's own observation. Therefore, the observation of individual 2 includes weak observation of previous people's observation with an eye of right before person.

1.1.4. d)

The 3 should accept in both case. Obviously, when 3 observed High and 2 accepted, this shows that majority has observed High in previous observations. When 3 observed low and 2 accepted, this can be looked as tricky condition but 2 has accepted can be understood as 1 and 2 both saw High so that is the reason why 2 accepted. Therefore, 3 should accept whatever he/she observed when 2 accepted.

1.1.5. e)

The cascade can be occur in this world. After first 2, 3 people, cascade will occur. As people after 3rd stage, they need to deduce the decision made by previous people, they will assume that their decision may be done as it is majority. Therefore, cascade can be occur.

1.2. Q4

1.2.1. a)

Simply, when first 2 people make wrong decision then whole network will guess wrong. The probability that 1 person guess wrong is $P(L|G) = P(H|B) = 1/3$. Therefore, to make whole R-cascade is wrong, only two person at the very first should make wrong decision which is happen with probability $1/3 * 1/3 = 1/9$

1.2.2. b)

With the observation provided to 10, the cascade can be collapsed. If 10 also observes High then, as 9 provided its observation is High, 10 will decide Accept the High. Otherwise when 10 observes Low then, 10 will Reject the Low.

1.2.3. c)

As the cascade is broken by showing the 9's signal to 10, the new cascade will start from 10. Because 11 knows that 10 has observed 9's signal, 11 will follow the cascade starting from 10. When 10 Accepts, 11 will also accept the signal and when 10 rejects, 11 will also reject the signal.

1.3. Q5

1.3.1. a)

Due to the "Herding" effect, even some committee members have thought that candidate B is the best choice, exceeding the cascade threshold toward candidate A made difficult to express their opinion that candidate B is better.

1.3.2. b)

Unanimous voting and blocking viewing others' choice may lead to fair decision making. As the previous committee meeting for the decision had some availability to observe other member's decision, the proposed method will block viewing other's decision which will make members to express their own opinion.

2. Network Book Exercise 19.8 - Q1, 2, 5, 6

2.1.Q1

2.1.1. a)

Node C, I, K will change to behavior A.

2.1.2. b)

Node G and J have 3 neighbors, so even node F or K become behavior A, they will not switch to behavior A due to the threshold $1/2$. Moreover, node D and H cannot changed to behavior A as G and H has blocked the spreading.

2.2.Q2

2.2.1. a)

Node E, F, H and I will change to behavior A.

As node F has 3 neighbors and 2 are behavior A, node F become behavior A. After node F becoming behavior A, this made node E to be behavior A ($2/4 > 2/5$) and made H and I also to be changed.

2.2.2. b)

Node G and J blocks the spreading. because both node has 3 neighbor nodes but only one neighbor D and F, respectively, is on behavior A. Therefore, they blocks the spreading.

2.2.3. c)

Wiring the node D and J will result all node to switch to behavior A. The new edge results node J to have 2 neighbors with behavior A which makes exceeding the threshold ($2/4 > 2/5$). This demolished the cluster that blocks the spreading. Therefore, node G switch to A ($2/3 > 2/5$) and also the node K ($2/3 > 2/5$).

2.3.Q5

Yes. However, the nodes will still make decision based on slightly different threshold.

For the case where opposite behavior get a payoff of 0, the threshold q to adopt behavior A from B can be obtained from following equation. $a \cdot p + 0 \cdot (1 - p) > 0 \cdot p + b \cdot (1 - p)$, where p is the ratio of neighbors wired to behavior A and $(1-p)$ is the neighbors wired to behavior B. From the equation p (which is the threshold q) should exceed $b/(a+b)$ to switch the behavior from B to A and switching from A to B can be done with threshold p is smaller than $b / (a+b)$ which means when a / $(a+b)$ fraction of B is on the neighbor.

At this point when the variable 'x' is introduced, the equations will be changed to $a \cdot p + x \cdot (1 - p) > x \cdot p + b \cdot (1 - p)$, where p remains ratio of the neighbors wired to behavior A and the solution for p from the equation fits when the node willing to change from behavior B to A. The threshold q to make the node switch behavior from B to A is $(b-x) / (a+b-2x)$ and to make the node switch behavior from A to B, at least $(a-x) / (a+b-2x)$ fraction of neighbors should take behavior B.

Here, when taking limit x to $0+$, both threshold from A to B and B to A will converge to the value that x is not introduced. Therefore, the decision make will not change.

2.4.Q6

Sample one student. The student will play the game if 2 of his/her friend play the game as there are 3 friends, one on the other floor and 2 on same floor. There are three cases that the student may play the game: [(2 friends on same floor play game), (1 friend (a) on same floor and 1 friend on the other floor play game), (1 friend (b) on same floor and 1 friend on the other floor play game)].

The probability that the sampled student may involved to the game is:

$$(5/10 * 4/9 * 1/2) + (5/10 * 5/9 * 1/2) + (5/10 * 5/9 * 1/2) = 7/18 < 1/2$$

As the probability is lower than $1/2$, the all 20 students in the group may adopt the game with very low probability.

Moreover, if the group of friends who play the game do not overlaps with the others do not play the game then the spreading will not occur.

Therefore, more information is needed to evaluate and although there is overlap between player group and non-player group, the spreading to all 20 students will occur in low probability.

3. COVID-19 Data Analysis

3.1. Dataset

The analysis is done on South Korea, United States, United Kingdom and worldwide.

The COVID-19 dataset includes ‘confirmed’, ‘recovered’ and ‘deaths’ for 816 days from initial report. Therefore, the time range is from 2020-01-22 to 2022-04-16. However, there is some outliers placed on about 500 day which make the modeling difficult. Therefore, only first 300 days will be involved in the analysis.

The population of each countries are give in following table.

Country	Population
South Korea	51 Millions
United States	329 Millions
United Kingdom	67 Millions
Global	7 Trillions

3.2. Models

The models used for the analysis are SIR, SEIR, and SEIRD. The ‘S’ stands for Susceptible, ‘E’ for Exposed, ‘I’ for Infected, ‘R’ for Recovered and ‘D’ for Deaths.

As the real data provides ‘confirmed’, ‘recovered’ and ‘deaths’, prediction for ‘I’ can be compared with the value ‘confirmed – recovered – deaths’.

The prediction’s error is measured with r2 score.

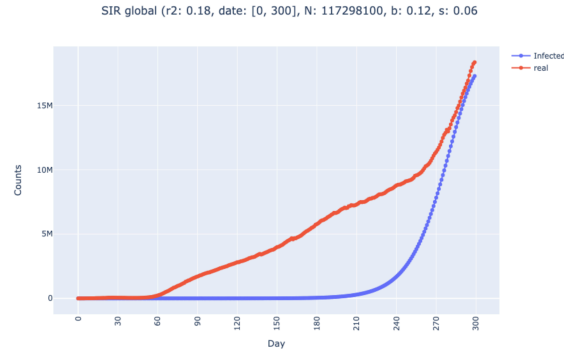
3.3. Analysis

In this section, manual hyper-parameter optimization is used.

3.3.1. World Wide

The SEIRD model shows best fit as it results highest r2 score. SIR and SEIR show similar performance. This result may be caused as most of countries did not collect information whether the people are exposed to COVID. Therefore, ‘E’ will not work as expected.

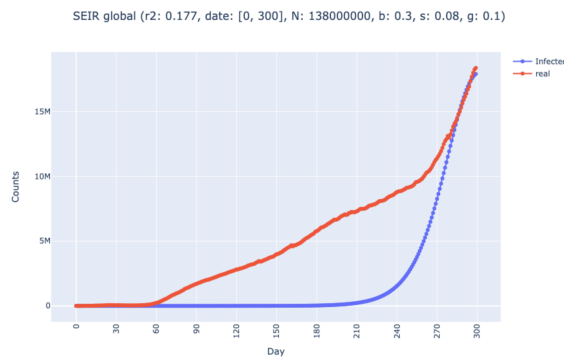
3.3.1.1. SIR



[Figure 1-1: SIR model for global]

The SIR model with 117M for initN, 0.12 for ‘beta’ and ‘0.06’ for ‘sigma’ results 0.18 on r2 score.

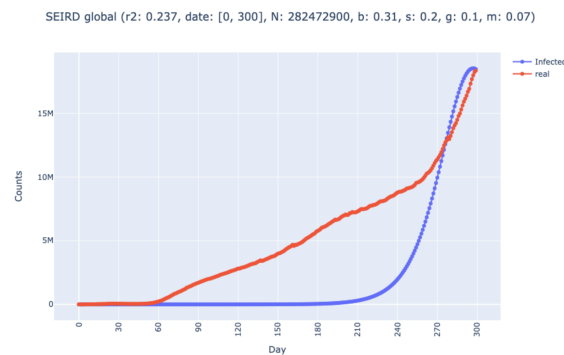
3.3.1.2. SEIR



[Figure 1-2: SEIR model for global]

The SEIR model with 138M for initN, 0.3 for ‘beta’, 0.08 for ‘sigma’ and 0.1 for ‘gamma’ results 0.177 on r2 score.

3.3.1.3. SEIRD



[Figure 1-3: SEIRD model for global]

The SEIRD model with 282M for initN, 0.31 for ‘beta’, 0.2 for ‘sigma’, 0.1 for ‘gamma’ and 0.07 for ‘mu’ results 0.237 on r2 score.

3.3.2. Korea

For the South Korea, due to the policy change, model fit are done on 0~150 days and 150~300 days separately. 150~300 days changed the baseline (0) as the real value on 150th day. This is setup for the convenience of model fit and analysis.

For all models initN is larger on 0~150 days than 150~300 days. This means that more population are involved to this pandemic during first 150days.

The SEIRD model shows best fit as it results highest r2 score. SEIR shows better performance on date range 150~300 days while 0~150 days are reduced. Because ‘E’ made the width of the model fit more close the the real data, the performance on 150~300 days has improved.

Ultimately, using SEIRD will take advantage on 150~300 days. Moreover, additional fine tune on SEIRD may be improved as much SIR achieved.

In conclusion, as multiple model used in different time series, model mixture can be considered to fit the mixture of the model into the real data for whole time range.

3.3.2.1. SIR



[Figure 2-1: SIR model for South Korea (left: 0~150 days, right: 150~300 days)]

The SIR model with 46k for initN, 0.4 for ‘beta’ and 0.19 for ‘sigma’ results 0.73 on 0~150 days. The SIR model with 16k for initN, 0.24 for ‘beta’ and 0.1 for ‘sigma’ results 0.577 on 150~300 days.

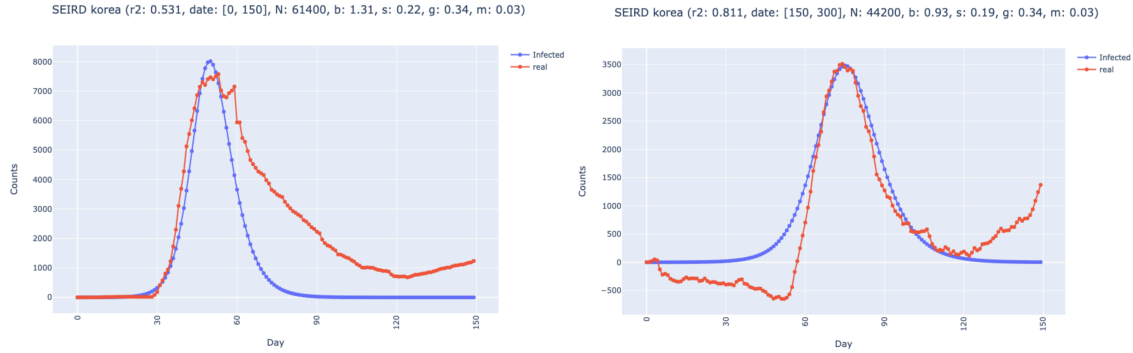
3.3.2.2. SEIR



[Figure 2-2: SEIR model for South Korea (left: 0~150 days, right: 150~300 days)]

The SEIR model with 55k for initN, 1.38 for 'beta', 0.19 for 'sigma' and 0.34 for 'gamma' results 0.534 on 0~150 days. The SEIR model with 41k for initN, 0.77 for 'beta', 0.26 for 'sigma' and 0.34 for 'gamma' results 0.782 on 150~300 days.

3.3.2.3. SEIRD



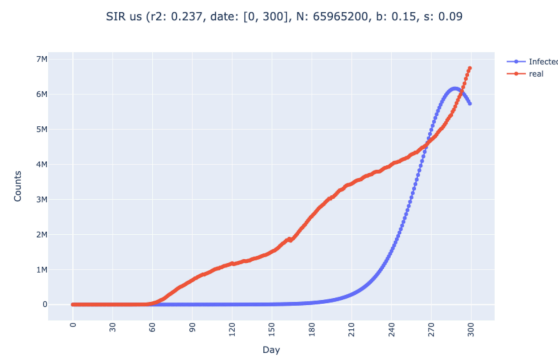
[Figure 2-3: SEIRD model for South Korea (left: 0~150 days, right: 150~300 days)]

The SEIRD model with 61k for initN, 1.31 for 'beta', 0.22 for 'sigma', 0.34 for 'gamma' and 0.03 for 'mu' results 0.531 on 0~150 days. The SEIRD model with 44k for initN, 0.93 for 'beta', 0.19 for 'sigma', 0.34 for 'gamma' and 0.03 for 'mu' results 0.811 on 150~300 days.

3.3.3. United States of America

For the US, as the nation did not collect the recovered and death properly. This data collection policy made the COVID-19 patient data on the US is difficult to figure out the actual infected people. Therefore, all model fit do not properly fit into the real data. The best performance model in this bad condition is SIR as E and D do not work properly with US dataset.

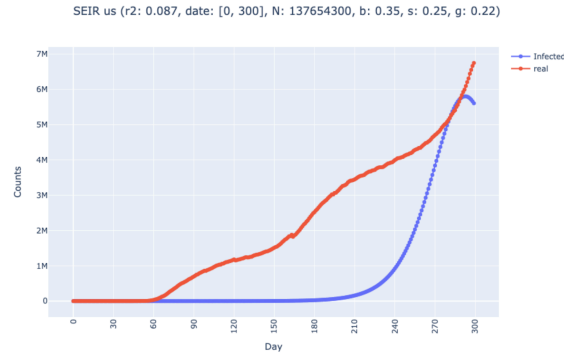
3.3.3.1. SIR



[Figure 3-1: SIR model for US]

The SIR model with 65M for initN, 0.15 for 'beta' and 0.09 for 'sigma' results 0.237.

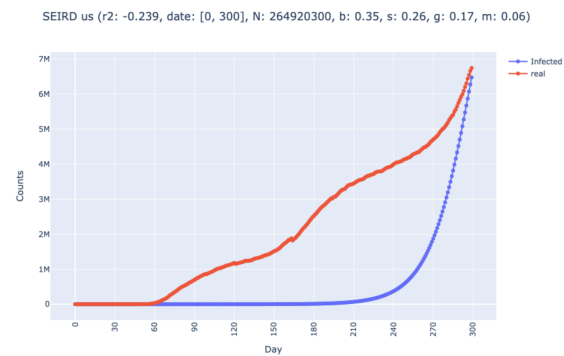
3.3.3.2. SEIR



[Figure 3-2: SEIR model for US]

The SEIR model with 137M for initN, 0.35 for ‘beta’, 0.25 for ‘sigma’ and 0.22 for ‘gamma’ results 0.087.

3.3.3.3. SEIRD



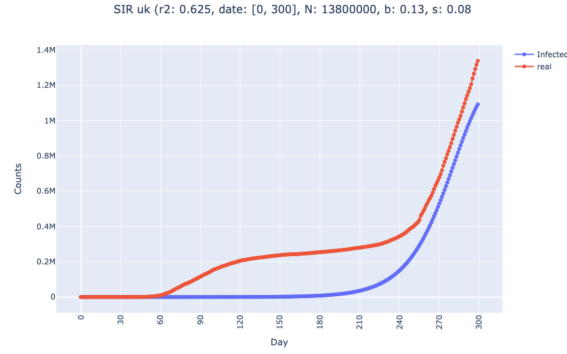
[Figure 3-3: SEIRD model for US]

The SEIRD model with 264M for initN, 0.35 for ‘beta’, 0.26 for ‘sigma’, 0.17 for ‘gamma’ and 0.06 for ‘mu’ results -0.239.

3.3.4. United Kingdom

For the UK, the nation collected confirmed, recovered and death cases quite well at the early stage of this pandemic. Therefore, the data used to fit the models quite worked well contrast to the US dataset. Although the UK collected data well, the spread of the COVID-19 was rapid than other nations after 200days. This phenomena caused all the model difficult to fit in to the dataset. Therefore, SIR showed the best r2 score and SEIRD and SEIR follows. However, as the hyper parameters are all tuned manually, there are some gap to improve the performance.

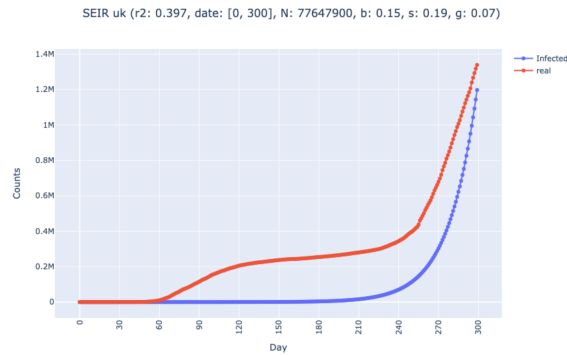
3.3.4.1. SIR



[Figure 4-1: SIR model for UK]

The SIR model with 13.8M for initN, 0.13 for ‘beta’ and 0.08 for ‘sigma’ results 0.625.

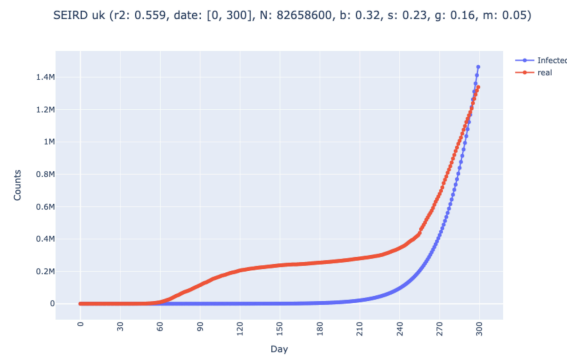
3.3.4.2. SEIR



[Figure 4-2: SEIR model for UK]

The SEIR model with 77.6M for initN, 0.15 for ‘beta’, 0.19 for ‘sigma’ and 0.07 for ‘gamma’ results 0.397.

3.3.4.3. SEIRD



[Figure 4-3: SEIRD model for UK]

The SEIRD model with 82.6M for initN, 0.32 for ‘beta’, 0.23 for ‘sigma’, 0.16 for ‘gamma’ and 0.05 for ‘mu’ results 0.559.

3.4. Hyper-Parameters

The hyper-parameters were tried to be optimized with grid search. All possible parameters are tested. However, the loss function, the r2 score, led the fit to some extreme cases such as Figure 5. The

red line drawn with SEIR obviously do not fit into the real data but the r^2 score showed 0.1 which is the best result in the grid search. This may be caused due to large gap and the small range of each parameter to be searched. The large gap and small range are used because of the computation complexity. The small gap and large range will find better parameter trivially but this causes expensive computation.



[Figure 5: (weird case) SEIR model for Korea]

Moreover, the COVID-19 dataset does not exactly follow SIR, SEIR, and SEIRD as most of nations tried various policies to slow down the spread and many people argued to enjoy the daily life as usual caused weird graph plot. In addition, after some period as many social groups argued that collecting the confirmed case is the invasion of the privacy and the increase of the cases were difficult to be collected properly, many countries stopped collecting the number of cases confirmed, recovered and deaths.

Therefore, further hyper parameter search is found that an extremely difficult work, autonomous hyper-parameters tuning is not handled in this paper.