Using Logistic Function and LSTM Networks to Predict the start date of USTC

Zhan Zhang, PB17000123

March 14, 2020

Abstract

The coronavirus disease 2019 (COVID-19) outbreak originating in Wuhan, Hubei province, China, coincided with chunyun, the period of mass migration for the annual Spring Festival. To contain its spread, China adopted unprecedented nationwide interventions on January 23 2020. These policies included large-scale quarantine, strict controls on travel and extensive monitoring of suspected cases. However, it is unknown whether these policies have had an impact on the epidemic. I sought to show how these control measures impacted the containment of the epidemic. I also used an artificial intelligence (AI) approach, trained on the 2003 SARS data, to predict the epidemic.

1 Introduction

Speaking at the COVID-19 media briefing, the WHO Director-General said: ”WHO has been assessing this outbreak around the clock and we are deeply concerned both by the alarming levels of spread and severity, and by the alarming levels of inaction.

We have therefore made the assessment that COVID-19 can be characterized as a pandemic.

Pandemic is not a word to use lightly or carelessly. It is a word that, if misused, can cause unreasonable fear, or unjustified acceptance that the fight is over, leading to unnecessary suffering and death.

Describing the situation as a pandemic does not change WHO's assessment of the threat posed by this virus. It doesn't change what WHO is doing, and it doesn't change what countries should do.

We have never before seen a pandemic sparked by a coronavirus. This is the first pandemic caused by a coronavirus.

And we have never before seen a pandemic that can be controlled, at the same time.”

2 Modified SEIR model

I modified the original SEIR-equation to account for a dynamic Susceptible [S] and Exposed [E] population state by introducing the move-in, In(t) and move-out, Out(t)

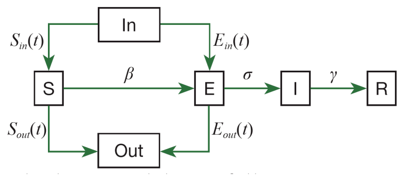


Figure 1: Model

parameters. Conceptually, the modified model is shown as: The base model is as follows:

$$\frac{dS(t)}{dt} = -\beta\frac{S(t)I(t)}{N}, \frac{dE(t)}{dt} = \beta\frac{S(t)I(t)}{N} - \sigma E(t), \frac{dI(t)}{dt} = \sigma E(t) - \gamma I(t), \frac{dR(t)}{dt} = \gamma I(t)$$

Here, we assume that latent [E] population is asymptomatic but infectious, and [I] refers to the symptomatic and infectious population. The incubation rate, σ is described as the rate by which the exposed individual develops symptoms.

2.1 Logistic Function

Since the model is so complicated, so we use logistic function to simplify it. A logistic function or logistic curve is a common ”S” shape (sigmoid curve), with equation:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

e = the natural logarithm base (also known as Euler's number),
x_0 = the x-value of the sigmoid's midpoint,
L = the curve's maximum value, and
k = the logistic growth rate or steepness of the curve.

Some may say people lives in different part, the must be simulated group by group, but I will say that, in large scale, the count will still present in the pattern of Logistic Function. So, We can aproximate our curve by using least squares method, here is what I approximated:

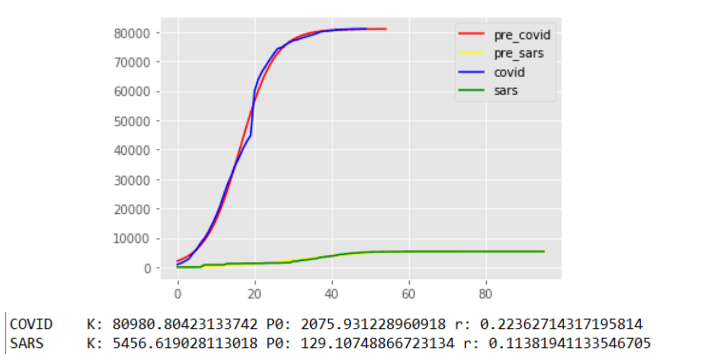


Figure 2: Logistic Function

2.2 LSTM Networks

Humans don't start their thinking from scratch every second. As you read this essay, you understand each word based on your understanding of previous words. You don't throw everything away and start thinking from scratch again. Your thoughts have persistence.

Traditional neural networks can't do this, and it seems like a major shortcoming. For example, imagine you want to classify what kind of event is happening at every point in a movie. It's unclear how a traditional neural network could use its reasoning about previous events in the film to inform later ones.

Recurrent neural networks address this issue. They are networks with loops in them, allowing information to persist.

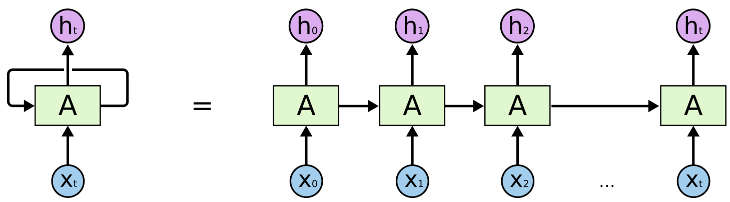
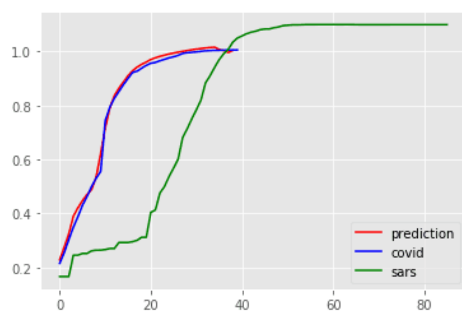


Figure 3: LSTM

Using the data of SARS_2013 and the approximated logistic function above, we can scale the dataset using the parameter in logistic function, and then predict the date by LSTM.I use privious 10 days to predict after 5 days, here comes my result:

```
------epoch 3000-------
loss:0.00003
min_loss:0.00001
test_loss:0.00035
Model saved in file net_params_confirmedCount.pkl
```




```
----------end-----------
Model saved in file net_params_confirmedCount.pkl
```

Figure 4: LSTM train confirmedCount

I also using LSTM to predict the suspected count. Since the dataset is small, I use privious 2 days to predict after 1 days, here comes my result:

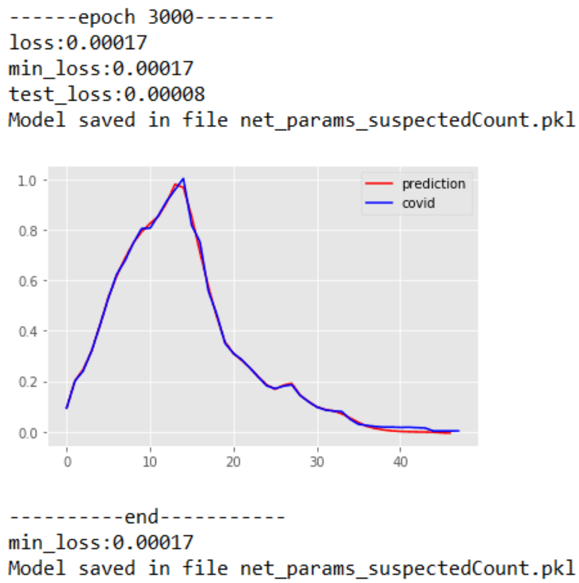


Figure 5: LSTM train suspectedCount

3 Prediction

Written down as a set of equations, LSTMs look pretty intimidating. Hopefully, walking through them step by step in this essay has made them a bit more approachable. Using these LSTM Networks, I predict what the count will be looked like in next 30 days:

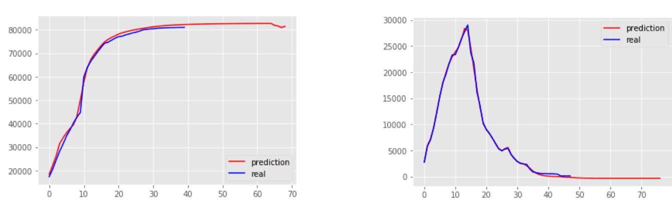


Figure 6: predict confirmedCount and suspectedCount

It is easy to find that, the confirmedcount grow very slowly in the next mounth, and the suspectedCount reach 0 at about 53 days(03/20/2020). Considering that the government's decision must be very careful, so they will observe the situation for a month. In conclusion, I thougth the start date of USTC is at 04/20/2020.

References

[1] Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Nanshan Zhong, 2020.

[2] https://github.com/BlankerL/DXY-COVID-19-Data

[3] https://www.kaggle.com/imdevskp/sars-outbreak-2003-complete-dataset

[4] https://b23.tv/av95917816