Predictions for Corona pandemic

Author and Reporter: An La

Outline

1. Objective
2. Exploring data in China
3. Modelling

1. Objectives

Answering the question when the corona pandemic will end:

- In the World
- In the US

1. Objectives

Last date of pandemic = L + x,

L is the date of last positive cases

x is the average of days of being infected. x in [14, 30].

Our objective becomes predicting L

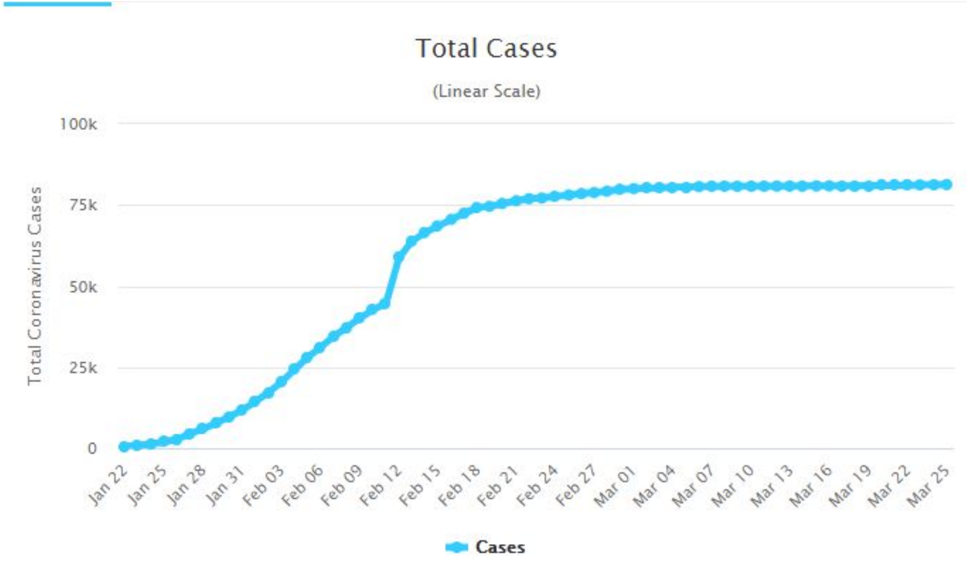
2. Exploring data in China

When L comes?

+ When people intervene successfully
+ When all people are infected

Evidence of L?

+ The number of **total cases** grow slightly, and close to the constant

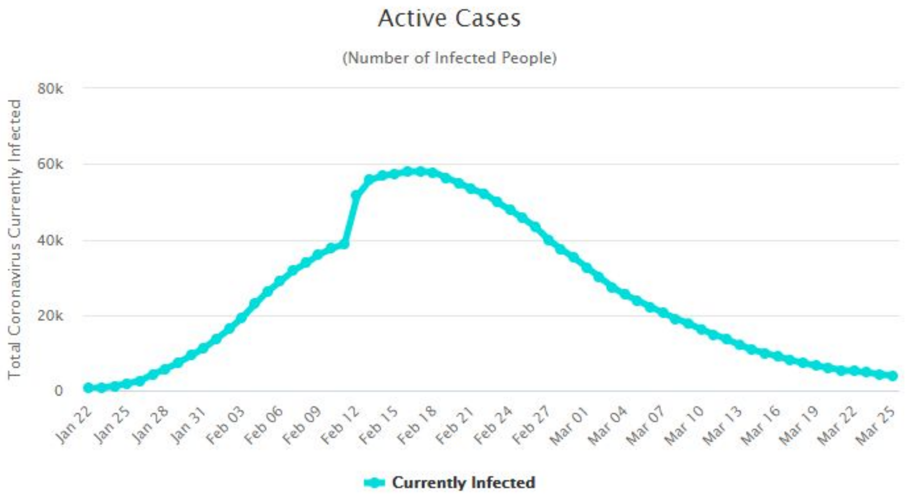


https://www.worldometers.info/coronavirus/country/china/

2. Exploring data in China

Evidence of L?

+ The number of **active cases** or **new cases** decrease strongly



https://www.worldometers.info/coronavirus/country/china/

2. Exploring data in China

*The researchers estimate that in the early stages of the Wuhan outbreak (from December 1, 2019 to January 25, 2020) each person infected with 2019-nCoV could have infected up to 2-3 other individuals on average, and that **the epidemic doubled in size every 6.4 days**.*

https://www.sciencedaily.com/releases/2020/01/200131114753.htm

$$2R^{x} = R^{x+6.4} \Rightarrow R = \sqrt[6.4]{2} = 1.1143867425959$$

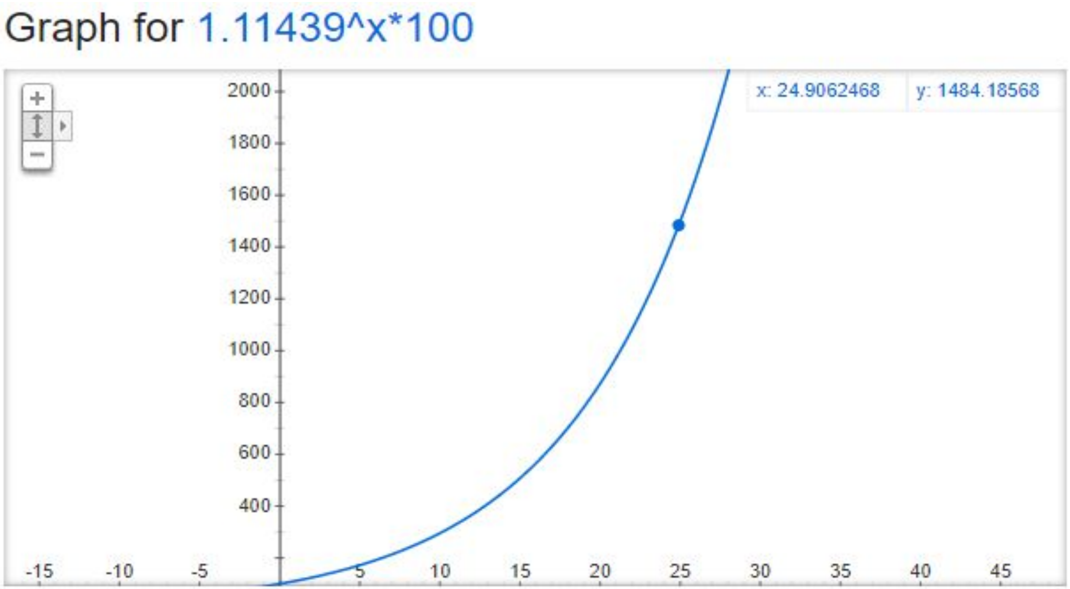
The number of people infected in day xth: $I \sqrt[6.4]{2}^{\,x}$

where I is the number of people infected in day 0

R is the speed of spreading virus. If R <= 1.0, we overcome the outbreak.

2. Exploring data in China

Suppose that there are 100 people infected in the day 0.
After 6.4 day, the number of people infected 1.11439^6.4 * 100 = 200



Graph for equation of epidemic size with R. x-axis is day, y-axis is the total infected cases

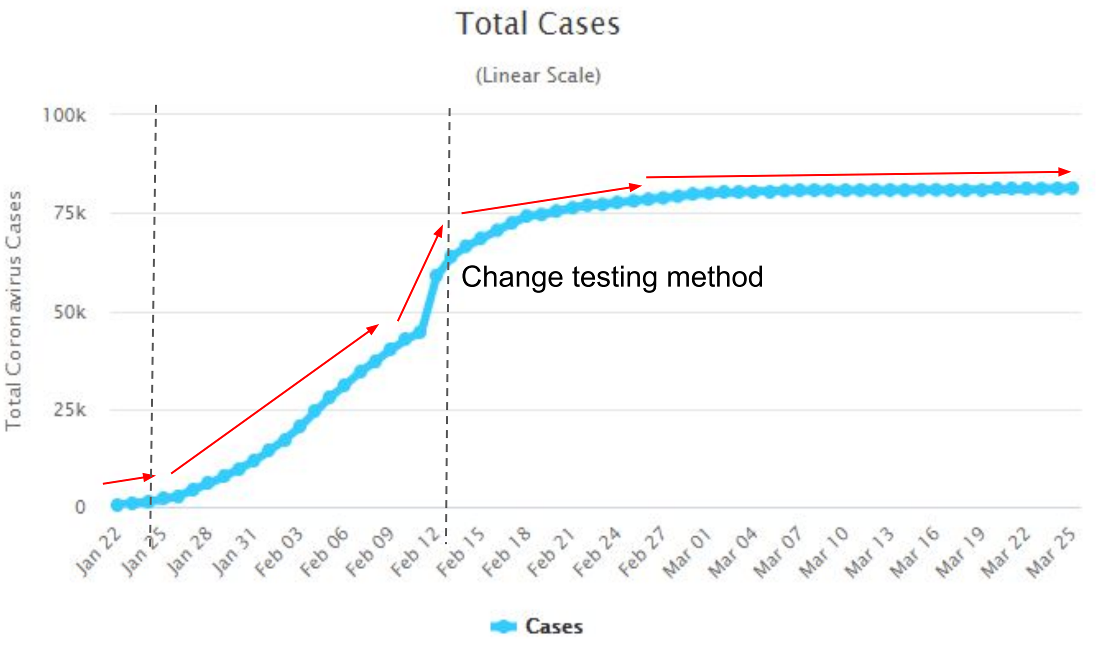
2. Exploring data in China

R is not always high. Before outbreaking and after being controlled, the speed is low.

R is also affected by testing method

Define 3 phases:

- Before outbreak
- Outbreak
- Ending



2. Exploring data in China

*A new study led by doctors at Tongji Medical College in Wuhan estimated that the virus infection rate – t**he average number of people infected with an infected person – was 3.68** in Wuhan before the closure began on January 23. However, **from February 2 to 18**, the number, which has to drop below 1 to stop the epidemic, **dropped to only 0.32**. The method, it seems, worked.*

https://www.news1.news/i/2020/03/not-just-isolation-what-do-western-countries-miss-by-the-method-that-worked-in-the-east.html

In case of Wuhan, without successful intervention, the ending phase does not come to soon and stop at around 80,000 people, but 11.08 million inhabitants (its population) are in risk.

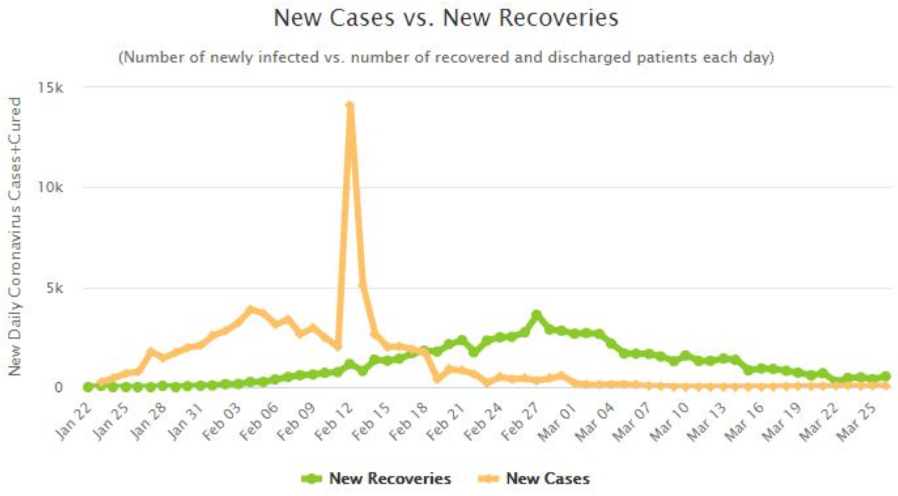
Let call C is the intervention, including policy of the government, changing in weather… Higher C leads to lower R.

2. Exploring data in China

However, in China, outbreaking lasts until Feb 18, not immediately stopped when intervention occurs. There are many relevant factors:

- Epidemiology: Virus survival in the environment, incubation, …
- **Health Care System**: the fact that hospitals are easy to be overloaded or become epicenter.
- ...

The ability of Health Care System is reflected by number of recoveries and death. The pandemic can be controlled if R is still large but fast recovery.



https://www.worldometers.info/coronavirus/country/china/

2. Exploring data in China

In conclusion

- L comes when no new cases: spreading is controlled or all people are infected.
- Spreading is controlled: R <= 1.0 or speed of recovery is high.

R is affected by

- Phase of virus spreading in community
- Intervention, especially controlling of human: This depends on policies nations, even states (in US). The most impacted attempts includes health care system, reducing activities and remaining distance between people, even the closure.

2. Exploring data in China

Therefore, to achieve the target:

- With world-level: answering the objective question for top countries affected by the virus in different continents, such as Italy, French and German; US, Canada and Brazil; Egypt and South Africa; Iran, Korean and Malaysia; Australia.
- With US-level: answering the question in detail of states.

3. Modelling

Summary of factors:

- **Location**: nations or states. Beside of human controlling, it is also an indicator for natural conditions (weather), population features (population, population density, popular age), other (transportation, business and culture activities). **Each location has independent model.**
- **Virus spreading phase:** we want to predict when the outbreak will end

Summary of observations: the number of

- Total cases (t)
- Active cases (a)
- New cases
- Recovered cases
- Death

All values are counting by days.

3. Modelling

- Predicting R from its history
- Predicting number of new cases, recovery and death.
- Models with population features
- Models of intervention

Predicting R from its history

Input: the data of new cases by days

Methods: predicting R of the next day from fixed interval of days.

Challenge: choosing length of the interval. If it is large, changing human controlling can make predicting wrong. It if is small, estimating has less accuracy.

Solution: Apply estimating between intervention actions, such as: school closing, airport closing, forcing inhabitants at home.

Challenge: R is not affected immediately when intervention occurs.

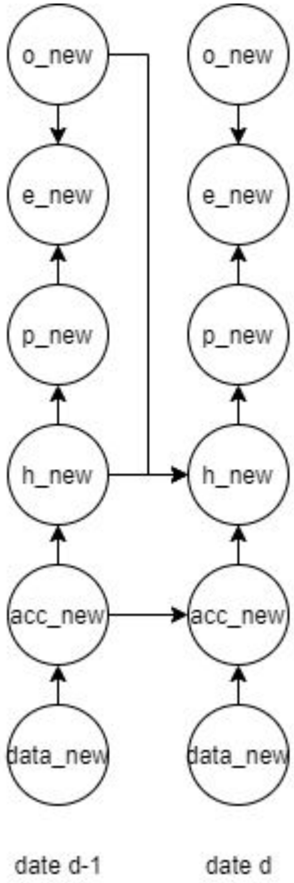
Solution: add to the interval a few days of incubation.

Predicting the number of new case

Model: Recurrent neural network

Data: Choosing interval of date as predicting R

Explaining nodes: data_new is number of new case in the date d, acc_new means accumulation, h_new is an activation function, p_new is results of prediction and o_new is output. Error is computed at e_new, comparing between prediction and output.



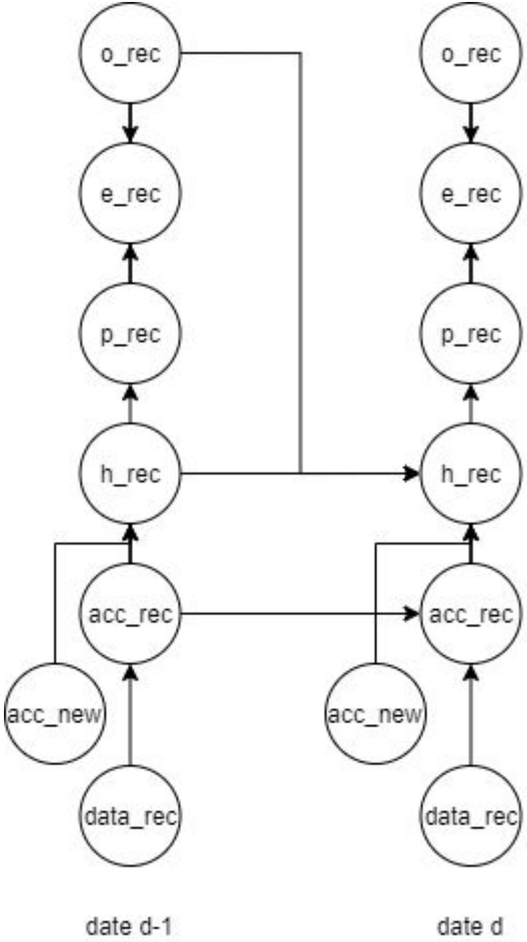
Predicting the number of recovered and dead

Model: Recurrent neural network

Data: Choosing interval of date as predicting R

Explaining nodes: similar to new, but h_rec also receives acc_new

The number of dead is similar to recovered



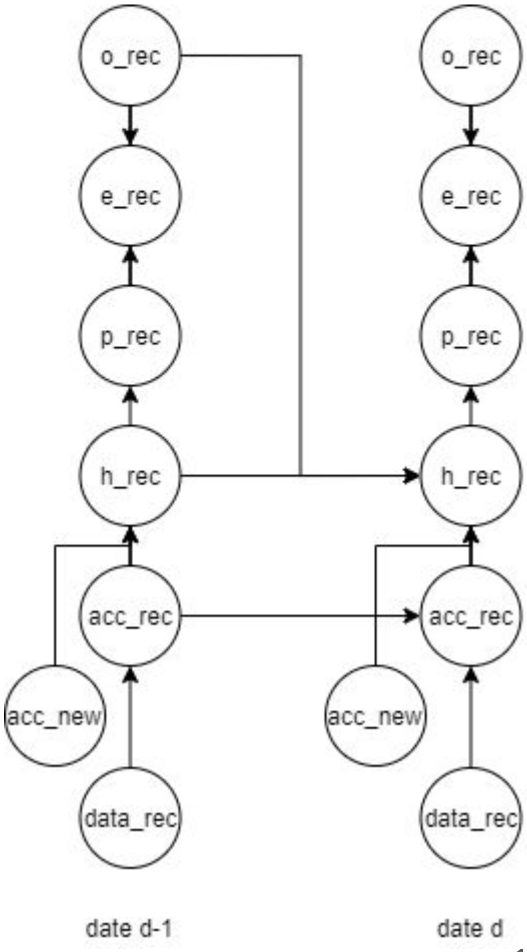
Predicting the number of recovered and dead

Model: Recurrent neural network

Data: Choosing interval of date as predicting R

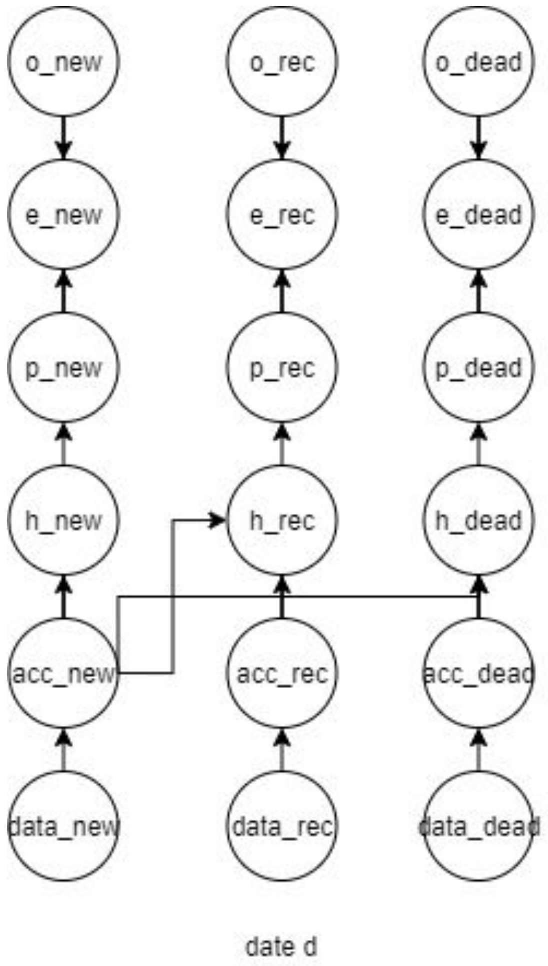
Explaining nodes: similar to new, but h_rec also receives acc_new

The number of dead is similar to recovered



Predicting the number of new case, recovered and dead

Training 3 models at the same time has better performance.



Models with population features

Both previous models can be improved by adding population features and we train all locations as them same time instead of each location has a independent model.

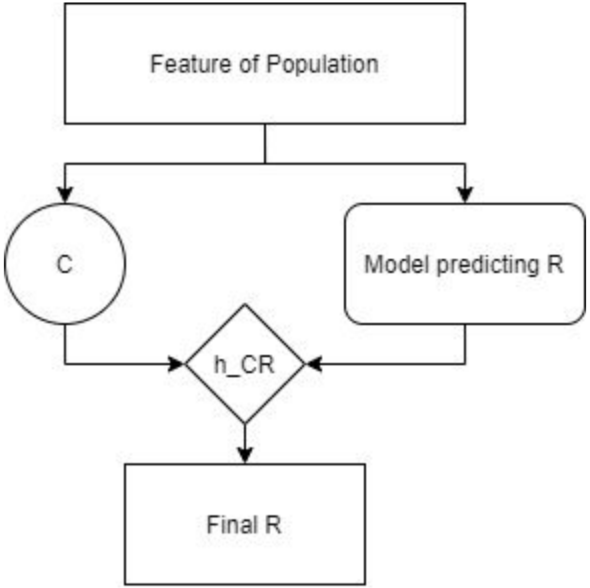
Population features: number of inhabitants, population density, age distribution, distribution of occupations

Models of intervention

In the section 2, we conclude that higher C (intervention) leads lower R.

- C can be computed by experts in epidemiology.
- This value is difficult to compute, so we use it as latent feature instead of calculating exact value.

In both case, C can be applied to model on the right, where h is a function exploits the relationship between C and R.



Modelling

Summary

- If predicted R is increasing, pandemic ends when
  - All people infected. From the population and average new cases by date, we can compute L.
  - Intervention makes the number of recovered increased. The relationship between speed of recovery and new cases should be analyzed.
- If predicted R is decreasing, pandemic ends when R <= 1.0. Calculating speed of changing R to compute L.

Modelling

Conclusion

- All values can be computed by estimating or apply recurrent neural network in time series.
- Some problem with data, such as the real number of positive is ambiguous, may not affected to the target.
- The most challenge for all cases is data size, especially recurrent neural network. Therefore, training multiple models at the same time can make them collaborative to each other, then increasing the accuracy.
- The model can be more precise with contribution from experts in epidemiology. Intervention should be explored more.