The Application of Logistic Model in COVID-19

Zhongyu Wang\textsuperscript{1st, 1}, Yuanhang Shi\textsuperscript{1st, 1}, Jiakun Xie\textsuperscript{1}

\textsuperscript{1}School of Electrical and Electronic Engineering, Shijiazhuang Tiedao University, China
\textsuperscript{2}School of Information Science and Technology, Shijiazhuang Tiedao University, China
wzy_stdu@126.com

\textbf{Abstract.} COVID-19 began to break out in China in early 2020, characterized by rapid transmission and a high fatality rate. In the face of the outbreak, it is particularly important and urgent to predict the number of infections in each region. Even now, some countries are still in the early stages of an outbreak. The existing traditional mechanistic models such as SIR have a high demand for data. For example, parameters such as exit rate need a large number of data to ensure the accuracy of the model; otherwise, the prediction effect is poor, while the models or algorithms related to artificial intelligence have a higher demand for data. Moreover, both of them have a large amount of work to solve, which has poor effect on the prediction of early epidemic prevention and control. The problem of epidemic situation prediction in the absence of data urgently needs to be solved. After data preprocessing, this paper used the Logistic model (the growth retardation model) to predict the number of infections of China. Meanwhile, error analysis and stability analysis were carried out, and the data of Italy were used for checking the universality of the model. The results show that the Logistic model is easy to be solved, has good stability and accuracy, and is suitable for early prediction of COVID-19.

1. Introduction

At the beginning of 2020, Wuhan Municipal Health Committee reported a case of clustering pneumonia of unknown cause, which caused extensive attention of health departments of China. COVID-19 was the cause of the outbreak in China in January 8, 2020. With the development of the research, Nanshan Zhong found that the epidemic was highly infectious, and its main transmission routes were respiratory tract infection and contact transmission. In just a few days, the virus swept through Hubei Province and even all over China. As of June 25, there were 85134 confirmed cases and 4648 deaths in COVID-19. Therefore, it is necessary to analyze the prediction of the epidemic situation.

The research method is to collect the cumulative number of confirmed cases, the cumulative number of cured cases, the cumulative number of dead cases and the cumulative number of suspected cases in different regions in different periods of time through network media [1], and then conduct statistical analysis and combine with relevant models to find the number of confirmed cases in different periods. On this basis, the correlation analysis and discussion of the predicted values are carried out. Mathematical modeling COVID-19 is important for the description, analysis, prediction and control of the spread of the epidemic.

This paper mainly analyzes the data provided by the National Health Commission of China, which is the basis for the prediction and analysis of the epidemic situation in China[2]. Literature[3][4] described the improved SIR mathematical model of COVID-19 transmission with different
characteristics; Literature [5] proposed the improved logistic growth model, which made a certain contribution to the prevention and control of SARS. Fuxia Xu et al proposed a universal growth retardation model to analyze the situation of SARS[6]. In literature [7], Nanshan Zhong’s team solved the improved SEIR model with the help of the deep learning algorithm LSTM, and predicted the epidemic trend after training according to the data about SARS. Zhixin Wang et al combined machine learning with SIR model to obtain comprehensive prevention and control measures [8]. Huaxiong Sheng et al used different models to predict the trend of epidemic situation in divisional periods [9]. Literature[10][13] proposed a modified SEIR model considering asymptomatic infections and incubation period. These studies have well predicted the trend of the relevant epidemic situation, but the method of using Logistic model completely in the above literature is only used in SARS, and does not predict the actual situation of COVID-19. The model solving or data preprocessing in literature[7][8] requires high training requirements for artificial neural network. When the training is insufficient, it is easy to form larger errors, and at the same time, it is difficult to solve the model to a certain extent. SIR related models are similar, lack of data in the early stage of the epidemic, so it is difficult to get good prediction results, and it is relatively cumbersome to solve the predictive problem accurately.

Based on the research methods of Yuelian Shi, Fuxia Xu and Huaxiong Sheng et al, this paper takes the epidemic situation in China as the research object, describes the virus transmission by Logistic model (growth retardation model), and introduced least square method to solve the parameter problem. Compared with the theoretical values and the actual values, the stability analysis and error analysis were carried out to verify the effectiveness of the model, and the data from Italy were used to verify the universality of the model.

2. Question hypothesis

2.1 Basic Hypothesis

(1) Ignore the birth and natural death of the population. Since the selected period is not long, it can be ignored.

(2) Ignore the recurrence after cure. Because the recurrence rate is very low after short-term cure, it can be ignored.

(3) Assuming that the epidemic statistics data the paper used are true and reliable;

(4) Ignoring the imported and exported cases;

(5) Assume that the total population of the whole country is a constant. Because of the large population base and the small number of deaths and returnees, so the resulting changes in the total population are negligible;

(6) Assume that, with the intervention of human society, the inherent growth rate of the total number of infections decreases approximately linearly with the increase of the total number of cases.

2.2 Definitions and Symbol Assumptions

| Symbol | Definition | Symbol Description |
|--------|------------|--------------------|
| $N$    | The total population |
| $I$    | The number of infections |
| $r$    | Inherent growth rate of number of patients |
| $\hat{I}$ | Estimates of the number of infections |
| $\sigma$ | Standard deviation |
| $t$    | Time |
3. Establishment of Model for Prediction

3.1 Establishment of Logistic model in COVID-19

It is found that the development mechanism of infectious diseases is similar to that of logistic model (differential equation model) of population, so it can be modeled by logistic model[6]

\[
\begin{align*}
    I'(t) &= rt(1 - IN^{-1}) \\
    I(0) &= 1
\end{align*}
\]

In the formula, \( t \) represents the days from the discovery of the first patient to the current date, \( N \) is the maximum number of infections in the country, and \( r \) is the inherent growth rate in the logistic model. \( r>0 \) in the model, the infectious disease is still in the spreading period. \( I(t) \) is the total number of confirmed cases on the \( t \) day

\[I(t) = N[1 + (N - 1)e^{-rt}]^{-1}\]  \hspace{1cm} (2)

The first case of infection in China appeared in December 8th, which was 63 days apart from February, and 37198 people were infected by the end of February 8th. The formula can be expressed as:

\[I(63) = 37198\] \hspace{1cm} (3)

The model with \( r \) of 0.16705 was calculated

\[I(t) = N[1 + (N - 1)e^{-0.16705t}]^{-1}\] \hspace{1cm} (4)

Virtually, the result obtained by this method isn’t very rigorous for lack of data.

4. Prediction of Epidemic Situation in China by Logistic Model

4.1 Data Preprocessing

The actual published data of China will be drawn into a scatter chart for rough observation. As shown in Figure 1.

![Patient number curve](image1.png)

Figure 1. Actual published data of infections of China

As shown in Figure 1, in order to make the prediction of the model more accurate, data preprocessing is necessary. Therefore, according to the growth of the number of infections in different periods, it is divided into two parts, and different regression models are used for data preprocessing.

4.1.1 Parabolic regression model

Because the detection rate of infections was relatively low in the early stage of the epidemic, and some patients were not detected in time, the epidemic data published by the National Health Commission of China was not consistent with the actual situation. Before the effective control in the early stage of the epidemic, the total number of infections was approximately a parabola, so the parabolic regression was used to fit the preliminary data for data preprocessing.

The relationship between the actual published number of infections and time was established as follows:
\[ I(t) = at^2 + bt + c \]  \hspace{1cm} (5)

By the least square method, we can solve:
\[
\begin{align*}
a &= 117.82 \\
b &= -457.06 \\
c &= 646.1579
\end{align*}
\]  \hspace{1cm} (6)

The regression model is:
\[ I(t) = 117.82t^2 - 457.06t + 646.1579 \]  \hspace{1cm} (7)

Draw the fitting curve and the actual data as shown in Figure 2.

4.1.2 Logarithmic Regression Model

Considering the later stage of the epidemic, the growth rate of the total number of infections actually gradually decreased, so the logarithmic regression model was used for data preprocessing.

The relationship between the number of infections and time was established as follows:
\[ I(t) = a \ln(t) + b \]  \hspace{1cm} (8)

By the least square method, we can solve:
\[
\begin{align*}
a &= 2937.41003498 \\
b &= 71920.14676856
\end{align*}
\]  \hspace{1cm} (9)

Namely, the regression model is:
\[ I(t) = 2937.41003498 \ln(t) + 71920.14676856 \]  \hspace{1cm} (10)

Draw the fitting curve and the actual data as shown in Figure 3.

4.2 Results of the Model

Logistic model was used to fit the pre-processed data, and a graph was drawn as shown in Figure 4.
Figure 4 indicates that for lack of effective control measures in the early stage of the epidemic, the growth rate of the total number of infections is relatively high, but in the later stage, under the effective control measures of the government, the growth rate of the total number of infections gradually decreases and finally almost zeros. The Logistic model can relatively accurately predict the number of infections with the intervention of human society by using the inherent growth rate $r$ obtained from the early short-term data and the fuzzy total population (the upper limit of the number of infections). Furthermore, more data can provide more accurate parameters, because parameters of logistic model can be solved by least square in the program.

4.3 Verification Based on Data of Italy
In order to check the accuracy and universality of the Logistic model in the prediction of the number of infections in COVID-19, this paper conducted an example verification based on the data of Italy. The previous outbreak in Italy was very serious[14]. Because of tough measures of the government, the assumption that the population is basically the same still holds.

According to the trend of the data, the data was first quadratic fitted for the former days and the fitting curve is shown in Figure 5:
Logistic model fitting was performed on the pre-processed data of Italy, and the obtained curve was shown in Figure 7:

As is shown in Figure 7, the daily growth rate shows a downward trend, consistent with the real situation. The predicted number of infections was in good agreement with the actual number. About 90 days after the selected date the curve was stabilized, and the number of infections hardly increases. The overall prediction effect is relatively good, indicating that the model has a certain universality and is not only targeted at specific data.

5. Analysis of Model Results

5.1 Error Analysis

In order to verify the accuracy of the model, the standard deviation, absolute and relative error will be calculated, and the standard deviation is calculated by comparing the overall data obtained by the model (China) with the actual published data (China). $D_1$ is the number of days for overall comparison, $I_i$ is the actual data, and $\hat{I}_i$ is the data obtained by the Logistic model (theoretical value):

$$\sigma_1 = \left[ \sum_{i=1}^{D_1} (I_i - \hat{I}_i)^2 D_1^{-1} \right]^{1/2} = 2258.5628$$

The standard deviation is calculated by comparing the previous data obtained by the model with the pre-processed data. $D_2$ is the days of overall comparison, $I_i$ is the pre-processed data, and $\hat{I}_i$ is the previous data obtained by the model (theoretical value):

$$\sigma_2 = \left[ \sum_{i=1}^{D_2} (I_i - \hat{I}_i)^2 D_2^{-1} \right]^{1/2} = 495.9549$$

The standard deviation is calculated by comparing the late data obtained by the model with the late
data after pre-processing. $D_3$ is the days of overall comparison, $I_i$ is the late data after pre-processing, and $\hat{I}_i$ is the late data obtained by the model (theoretical value):

$$\sigma_3 = \left[ \left( \sum_{i=1}^{D_3} (I_i - \hat{I}_i)^2 \right) D_3^{-1} \right]^{1/2} = 273.9455$$  

(13)

The standard deviation of the actual published data and the theoretical data of the Logistic model is larger due to reasons such as the patient's failure to be detected in time, while the standard deviation of the pre-processed data is significantly lower.

The daily relative and absolute error are as follows:

![Figure 8. Absolute error](image)

![Figure 9. Relative error](image)

In Figure 8, a serious fault appeared near the 30th day, which was caused by the fact that China began to change the means of screening, nucleic acid screening was adopted, and then the number of people detected with the disease increased sharply, which also affected the relative error nearby.

In Figure 9, the relative error is obviously large at the beginning, which is caused by a large number of infections in the initial stage who have not been detected, while the total number of infections detected at this time is relatively small, resulting in a large relative error.

To sum up, the error conforms to the objective law, and the model has certain accuracy and objectivity.

5.2 Stability analysis

In order to test the stability of the Logistic model, stability analysis should be carried out, which is whether the final convergence results of the established Logistic model for COVID-19 will be affected when the known quantity changes.

The total population (the upper limit of the number of infections) is 1.4 billion, and the inherent growth rate is 0.16705. The results of the model are as follows:

![Figure 10. Prediction results of the model ($N = 1.4$ billion; $r = 0.16705$)](image)

(1) In order to test the stability of the model, the upper limit of the number of infections is changed to 1.3 billion and 1.5 billion respectively. And the inherent growth rate remains unchanged at 0.16705. The solution results of the model are as follows:
As can be seen from Figure 11, when the upper limit of the number of infections was reduced to 1.3 billion or the upper limit of the number of infections was increased to 1.5 billion, the predictive fitting curve of the Logistic model remained converged. As long as the variation range of parameters is not too large, the error of N has little effect on the model result because the parameters can be recalculated by least square method inside the program used to solve the model.

(2) When the inherent growth rate was reduced from 0.16705 to 0.1 and increased to 0.3 respectively, and the upper limit of the infections remained unchanged at 1.4 billion, the model was solved as follows:

As can be seen in Figure 12, whether the inherent growth rate drops or rises within a certain range, the Logistic model still keeps convergence, and the final convergence value changes slightly. The main reason is that the program used to solve the model can determine these parameters by least square again. Therefore, it can be seen that this model is very stable when used to predict the number of infections.

To sum up, when the initial input value of Logistic model changed within a certain range, the final predicted results still converged despite slight changes, and the number of infections fluctuated within a small range. This fully indicates that the Logistic model has sufficient stability in the prediction of the number of confirmed COVID-19 cases, which is also its significant advantage.

6. Conclusion and prospect
The model used in this paper can well reflect the initial outbreak stage of the epidemic, and the prediction of post-peak has considerable accuracy[15]. Due to the structure of the model itself, compared to the general mechanistic models (such as SIR model, SEIR model) and network dynamics models (such as cellular automaton model, artificial neural network model) this model is much easier to program to solve, and the workload and difficulty of the prediction work were greatly reduced, making it possible to save a lot of precious time in the stage of selecting epidemic control measures in the early stage of the epidemic. The Logistic model does not need to consider the transmission mode of the virus, while traditional models such as SIR model needs to assume that the virus can only be transmitted from person to person, otherwise more influencing factors will be introduced. The established model has relatively
good stability. Even when the fluctuation of related parameters is large, the impact on the model results is still relatively small. In the case of insufficient data, the accuracy of the results in a short time is guaranteed to some extent, and prediction in the early stages of an outbreak has a huge advantage.

Actually, the epidemic occurred during the Spring Festival of China, the population migration could not be ignored. If the prediction of the large cities of China except Wuhan was made, the basic assumption that the total population in the region was constant or approximately constant would be invalid[16][17]. When a large number of relevant data is known, the accuracy of the theoretical data obtained by the model is lower than that of the artificial neural network model after extensive and targeted training, because in fact \( r \) is not strictly linearly decreasing with the increase of \( I \). When the data is insufficient, the long-term prediction effect is not so good. At the same time, for some countries or regions that have been subjectively inactive for a long time, the assumption that \( r \) decreases linearly with the increase of \( I \) approximately is invalid, and the prediction results are poor in a certain period.

The total number of infections in China increased significantly in March and April due to imported cases from abroad, which had a great impact on China's epidemic prevention work[18]. In the future epidemic prevention and control work of China, the prevention and control of imported cases from abroad will remain the top priority. In further research, a prediction that takes into account imported cases from abroad should be given. Besides, the distribution of vaccines should be considered in further research.

Reference

[1] Qian Li, Biao Tang, Nicola Luigi Bragazzi, Yanni Xiao, Jianhong Wu 2020 "Modeling the impact of mass influenza vaccination and public health interventions on COVID-19 epidemics with limited detection capability", *Mathematical Biosciences*

[2] National Health Commission of China 2020 *Outbreak announcement*. http://www.nhc.gov.cn/xcs/yqtb/list_gzbd_3.shtml. (in Chinese)

[3] Renjie Zhu, Shihao Tang, Tongtong Liu, Yan Guo, Shanshan Dong, Ying Cheng and Tielin Yang 2020 COVID-19 epidemic prediction based on improved SIR model and the impact of prevention and control on epidemic development. *Journal of Shanxi Normal University (Nature Science Edition) vol 48* p33-38 (in Chinese)

[4] Zi Yu, Guiqing Zhang, Qingzhen Liu and Zhongquan Lv 2020 The Outbreak Assessment and Prediction of COVID-19 Based on Time-varying SIR Model *Journal of University of Electronic Science and Technology of China* vol 49 p357-361 (in Chinese)

[5] Yulian Shi and Can Li 2019 Global dynamics of a class of SIRI infectious disease model with logistic growth and saturation incidence *Journal of Mathematics in Practice and Theory* vol 49 p299-307 (in Chinese)

[6] Fuxia Xu, Yongquan Dong and Dianshen Li 2005 Propagation model of SARS *College Mathematics* vol 4 p1-6 (in Chinese)

[7] Yang Z, Zeng Z and Wang K 2020 Modified SEIR and AI prediction of the Epidemic of CoVID-19 in China Under public Health Interventions *Journal of Thoracic Disease* vol 12 p165-174.

[8] Zhijun Wang, Zhi Liu and Zhaojun Liu 2020 COVID-19 analysis and forecast based on machine learning *Journal of Biomedical Engineering Research* vol 39 p1-5 (in Chinese)

[9] Huaxiong Sheng, Lin Wu and Changliang Xiao 2020 Modeling analysis and prediction on NCP epidemic transmission *Journal of System Simulation* vol 32 p759-766 (in Chinese)

[10] Hui Geng, Anding Xu, Xiaoyan Wu, Yong Zhang, Xiaomei Yin, Mao Ma and Jun Lv 2020 Analysis of the role of current prevention and control measures in the epidemic of corona virus disease 2019 based on SEIR model *Journal of Jinan University (Natural Science & Medicine Edition)* vol 41 p175-180. (in Chinese)

[11] Yongyue Wei, Zhenzhen Lu, ZhiCheng Du, Zhijie Zhang, Yang Zhao, Sipeng Shen, Bo Wang, YuanTao Hao and Feng Chen 2020 COVID-19 trend analysis based on improved SEIR–(+$CAQ$) infectious disease dynamics model Chinese Journal of Epidemiology vol 4 p470-475 (in Chinese)
[12] Shengli Cao, Peihua Feng and Pengpeng Shi 2020 Study on the epidemic of COVID-19 in Hubei province by modified SEIR model *Journal of Zhejiang University* (Medical Sciences) vol 49 p178-184. (in Chinese)

[13] TANG B, WANG X and LI Q 2020 Estimation of the transmission risk of 2019-ncov and its implications for public health interventions *J Clin Med* vol 9 p 462.

[14] Remuzzi A and Remuzzi G 2020 COVID-19 and Italy: what next? *The Lancet* 2020 vol 41 p7-12.

[15] Hao Li, Deaguang Dua, Xueqiang Tao, En Chen and Shutian Gao 2020 Review on infection disease dynamics models and their application in COVID-19 epidemic simulation and prediction *Chinese Medical Equipment Journal* vol 41 p7-12. (in Chinese)

[16] Xiaoke Xu, Cheng Wen, Guangyao Zhang, Haochen Sun, Bo Liu and Xianwen Wang 2020 The geographical destination distribution and effect of outflow population of Wuhan when the outbreak of COVID-19 *Journal of University of Electronic Science and Technology of China* vol 49 p324-329 (in Chinese)

[17] Jianxiong Hu, Guanhao He, Tao Liu, Jianpeng Xiao, Zuhua Rong, Lingchuan Guo, Weilin Zeng, Deixin Gong, Lihua Yin, Donghua Wang, Lilian Zeng and Wenjun Ma 2020 Export risk assessment of Hubei Province in the early stage of COVID-19 epidemic *Chinese Journal of Preventive Medicine* p362-366 (in Chinese)

[18] The National Health Commission held a teleconference to make arrangements for the next stage of prevention and control in accordance with the law, scientific prevention and control, and precise prevention and control *China health law* 2020 (in Chinese)