Evaluation and Prediction of COVID-19 Based on Time-Varying SIR Model

Liu Yanzhe and Liu Bingxiang[(✉)]

Jingdezhen Ceramic Institute, Jingdezhen 333403, China
`lbxl966@l63.com`

Abstract. This article based on the law of the COVID-19 epidemic situation until May 6 improves the SIR model. Through the reverse solution of the parameters in the model, the parameters are modeled and predicted, and the parameters that change with time are obtained. Compared with setting fixed parameters, the accuracy of the model is greatly improved. Using the improved model to analyze the COVID-19 epidemic situation and study the virus spreading trend in different countries. The results show that the improved model is basically reliable in predicting the development trend of the COVID-19 epidemic; the epidemic in Italy will basically end in July; the development trend in Britain and America is similar, and the inflection point is expected to appear in mid-June. The results of the study confirm the effectiveness of measures such as reducing the movement of people and providing medical assistance to the epidemic-stricken areas, and provide reference for subsequent epidemic prevention and control.

Keywords: COVID-19 · SIR model · Prediction

1 Introduction

Since the outbreak of COVID-19, the number of infected people worldwide has reached more than 3 million, and almost all countries have suffered huge losses. Many countries have taken measures to shut down various organizations to reduce population contact to prevent the spread of the virus, but these measures have a huge impact on the economy. Therefore, it is necessary to evaluate the development of the epidemic and provide a reference for policy formulation.

Most of the current studies are based on the SIR model to assess and predict the development of the epidemic [1–3], but in these studies, the most critical parameters of the SIR model are set estimates mostly. Since the spread of the epidemic will change with policy changes and the number of patients, this treatment can cause large errors. In this paper, based on the epidemic development rules, the equations are fitted to the parameters, and the SIR model is established using the parameters that change over time, finally, the model is verified, and this model is used to make predictions and analysis of the epidemic development.

2   Model Theory

The SIR model is a classic infectious disease dynamic model, this model is established by Kermack and McKendrick in 1927 [4]. Based on the SIR model and the characteristics of the COVID-19 epidemic, this article makes the following basic assumptions:

1) Since the change rate of epidemic situation changes over time is much more significant than that of births and deaths over time, and most countries have adopted strict immigration control measures, the total population change in a country is very small, so it is assumed that the total population in a warehouse keeps is a constant.

2) Infection rate coefficient = average number of patients in daily contact × probability of infection of susceptible persons after contact with patients, the average number of patients in contact is closely related to the government's prevention and control measures, people's awareness of isolation, etc. The average contact rate is high at the beginning of the outbreak. After a period of development, the average contact rate will gradually decrease after the outbreak is paid attention to. Therefore, the infection rate coefficient should be set as a variable parameter that changes with time.

3) The removal rate is related to the cure rate and mortality rate, however, because the number of dead patients accounts for a small proportion of the total number of patients, the removal rate and cure rate have a greater correlation. In the early stage of the outbreak, due to insufficient knowledge of the virus, the cure rate of patients is low. After accumulating a large amount of treatment experience, the patient's cure rate will rise and the removal rate will also rise. Therefore, the removal rate coefficient should also be set to change with time.

4) The data is in units of days and does not consider continuous changes.

Based on the above assumptions, this paper constructs the following balance equation:

$$\begin{cases} \frac{dS}{dt} = -S \times \frac{I}{N} \times \beta \\ \frac{dI}{dt} = S \times \frac{I}{N} \times \beta - \gamma \times I \\ \frac{dR}{dt} = \gamma \times I \end{cases} \tag{1}$$

$$\beta = a \times \ln(b \times t) + c \tag{2}$$

$$\gamma = k \times t + m \tag{3}$$

Among them, S, I, R denote susceptible persons, patients, and removed persons. β indicates the probability that a susceptible group will be infected after being exposed to infected crowd. γ is the coefficient of removal rate, indicating the probability of the patient being removed (dead or cured). At the same time, according to the above formula, the solving formula of basic reproduction number (R0) can be derived:

$$R0 = \frac{\beta}{N} \times \frac{1}{\gamma} \times S_0 \tag{4}$$

In this paper, based on the existing data, the parameter values in the model are solved in reverse. According to the obtained parameter data, a parameter model is constructed and trained. The parameter model obtained is used to predict the time-varying parameter data, then the time-varying parameter is used to construct a time-varying SIR model. When solving parameters in reverse, it can be solved directly, or optimization algorithms such as ant colony algorithm and genetic algorithm can be used to find the optimal parameter value of the model. After a comparative experiment, the results obtained in multiple ways are the same. Machine learning methods can be used as alternatives for situations where direct solutions are not possible.

This article attempts to use four methods of linear regression, polynomial fitting, exponential smoothing, and LSTM [5, 6] to establish the model of parameters β and γ. Since the parameter changes are more dramatic in the early stage of the epidemic and tend to be gentle in the later stage, the LSTM algorithm will easily cause the gradient to disappear. For the same reason, the use of exponential smoothing will also cause large errors. According to the law of epidemic development, the logarithmic function is finally used to establish the model of parameter β, and the univariate linear regression is used to establish the model of parameter γ.

3   Prediction Experiment

This article obtained COVID-19 data from four countries including China, Italy, Britain and America, including the number of patients, the cumulative number of deaths, and the cumulative number of cures, and calculate the number of existing patients in each country. Use the improved model to process data, verify the accuracy of the model, and predict the development of the epidemic.

3.1   Model Verification

Since the epidemic in China has basically ended, and the epidemic in Italy has also passed its inflection point, the accuracy of the model can be verified by predicting the development of the epidemic in China and Italy.

China's data began on February 5, 2020, with 21 training data, 70 prediction data, and 70 verification data. The fitting equation of β obtained by training is: y = −0.069 11675658517227 * ln (5.086007728270503 * x) + 0.3247851719047432, and the fitting equation of γ is: y = 0.0016630688923199239 * x + 0.014222881482399054; The Italy's data began on March 11, 2020, with 25 training data, 80 prediction data, and 30 verification data. The fitting equation of β obtained by training is: y = −0.059 88265296840833 * ln (5.095313307692628 * x) + 0.35327576998120064, and the fitting equation of γ is: y = 1.1499257648340639e^{-10} * x + 0.03115033914997176. The parameter fitting curves is shown in Fig. 1:
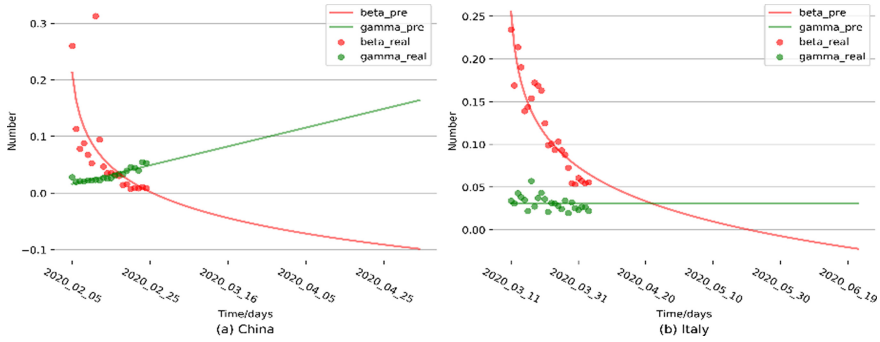
Fig. 1. Parameter fitting curves of China and Italy

The predicted parameters are used to build the model, and the resulting epidemic development curve is shown in Fig. 2. It can be seen from the fitting degree of the curve in the figure that the model has made a very good prediction on the development of the epidemic in China and Italy, and accurately predicts the development trend of the epidemic, the inflection point time and the number of infections. The results show that the Italian epidemic will basically end in July 2020. Data from China and Italy prove that the model has reliable effects.
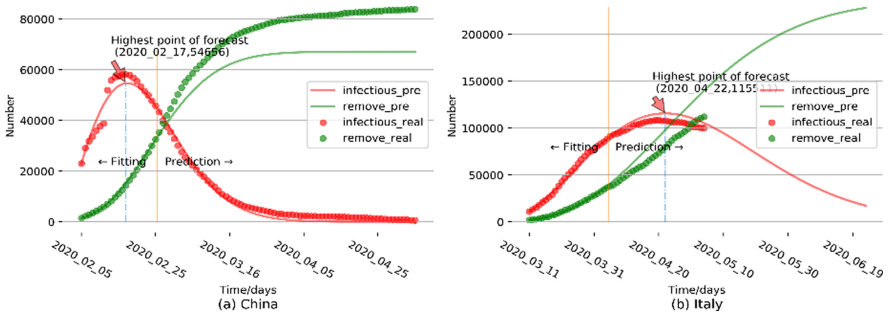

Fig. 2. Parameter fitting curves of China and Italy

3.2   Prediction

Apply the model to Britain and America to predict the development trend and inflection point of the epidemic in both countries.

Britain's data began on April 4, 2020, with 31 training data, 200 prediction data. The fitting equation of β obtained by training is: y = −0.026034231995124547 * ln (5.045657485354332 * x) + 0.16468626328436686, and the fitting equation of γ is: y = 1.009298676638269e^{-10} * x + 0.009474665494227158; American data began on April 5, 2020, with 30 training data, 200 prediction data. The fitting equation of β

obtained by training is: y = −0.0217276978288313 * ln (5.036525868374025 * x) + 0.13948441755802096, and the fitting equation of γ is: y = 1.3701824743687221e^{-08} * x + 0.012148954330530802. The parameter fitting image is shown in Fig. 3:
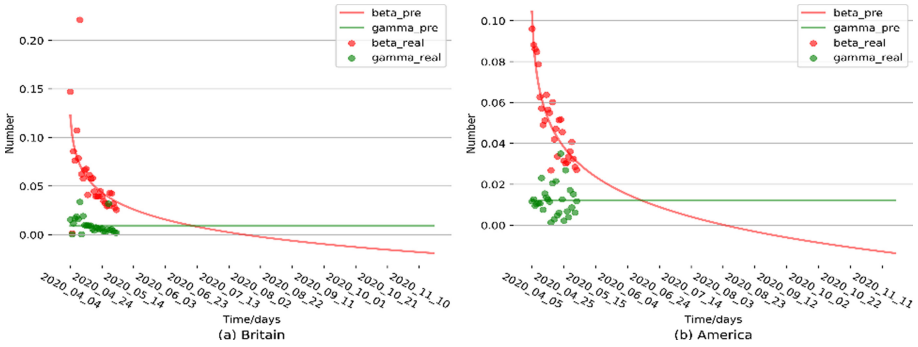

Fig. 3. Parameter fitting curves of Britain and America

Using the predicted parameters to build a model, the epidemic development curve is shown in Fig. 4. The model predicts that the inflection point of Britain epidemic will be June 19, when the number of patients on that day is 257673, and the end of Britain epidemic will be in December 2020. The predicted America epidemic inflection point is June 13, when the number of patients on that day is 1312227, and the end of the American epidemic is also about December 2020.
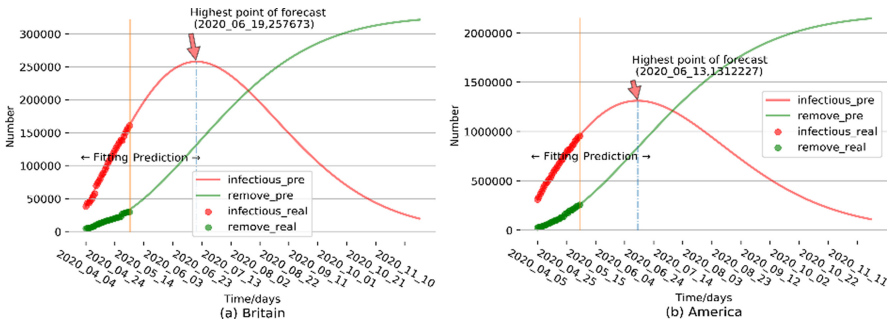

Fig. 4. The epidemic development curve of Britain and America

3.3   Analysis

The four countries have similar fitting equations for the parameter β, and the fit is very good. It shows that the infection rate coefficient of the epidemic has a fixed development law, and at the same time, it will cause some differences due to different national policies. The equations in China and Italy are similar, and the equations in Britain and

America are similar. It is linked to that China and Italy have adopted stricter prevention and control measures, proving that the prevention and control measures adopted by the government on the epidemic have a good impact on the development of the epidemic.

However, except for the model on China that has a good fitting effect on the parameter γ, none of the other three models can accurately obtain the fitting curve of γ. After the outbreak, China quickly mobilized national resources to support severely affected areas, so the cure rate has been significantly improved. However, the other three countries did not receive timely assistance after the nationwide epidemic broke out, so there is no obvious trend in the remove rate. At the same time, the fitting equation of γ shows that even without assistance, Italy's remove rate is still higher than that of Britain and America, proving that Italy's response measures have played a positive role.

According to the calculation formula of *R0*, the change curve of *R0* of four countries can be obtained. In order to make the comparison results more intuitive, the forecast days of Britain and America are reduced to 100 days, and the training days are unchanged (ensure the fitting equation remains unchanged), and the drawn curve is shown in Fig. 5.
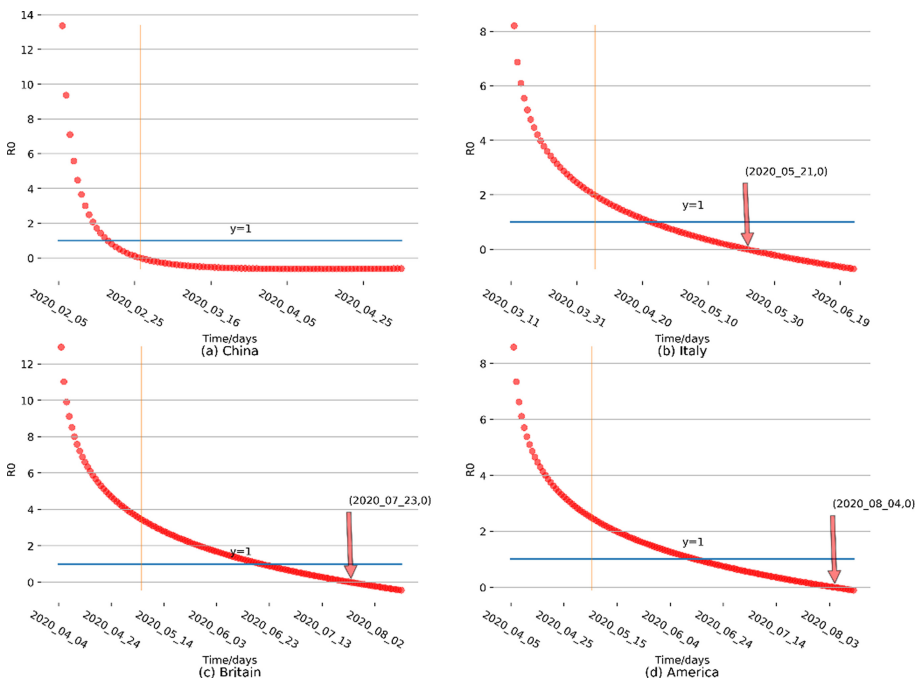

Fig. 5. R0 curve

It can be seen from the figure that China's R0 decreases the fastest, and it takes less than 20 days to reduce the R0 from 8 to 0. Italy's rate of decrease is also very fast, it is estimated that it takes about 40 days to reduce the R0 from 8 to 1. And Britain and America are estimated to take 70 days to reduce the R0 value from 8 to 1.

When the R0 is less than 1, the epidemic will no longer continue to spread, and it is not suitable for carrying out production activities before that. If judged according to the condition of R0 < 1, Italy, Britain and America can resume work on April 22, June 19 and June 13 respectively. However, when China started to resume work at the end of February, the R0 value was already less than 0. If judged according to this condition, Italy, Britain and America can resume work on May 21, July 23 and August 4.

In addition, a large number of studies on the R0 of the epidemic, it shows that the R0 of COVID-19 is between 2–8 [7–10].According to the predicted R0 curve analysis, if it is assumed that the initial stage with less human intervention is 20 days, the R0 in this stage is mainly distributed between 3 and 8, so the study believes that the R0 should not be less than 3, or even may possible be higher. The study also provides a reference for the estimation of R0 from the side.

4   Conclusion

Based on the SIR model, this paper establishes a fitting equation for the model parameters, estimates the parameters, and predicts the development trend of the epidemic according to the estimated parameters. Based on the data of COVID-19 diagnosis, death, and cure cases in 4 countries including China, Italy, Britain and America, the improved model is used to simulate and predict the data from the four countries. The results show that the improved SIR model can predict the epidemic trend reliably; the government's prevention and control measures can reduce the epidemic's infection rate coefficient and reduce the epidemic's spread rate; Assistance to the medical system in the outbreak area helps to increase the removal rate of patients; the average removal rate of Britain and America are similar, and both are significantly lower than the removal rate of Italy; The inflection point of the epidemic in Britain is on June 19, at this time the number of patients is 257673, and the end of the outbreak is approximately December 2020; the inflection point of the epidemic in America is on June 13, at this time, the number of patients is 1312227, and the end of the outbreak is also at December 2020; According to the conditions for resumption of work in China, people in Italy, Britain and America can return to work on May 21, July 23 and August 4 at the earliest; According to the R0 curve simulated in this study, the R0 of the COVID-19 epidemic should be at least 3 or more. Research results confirm that measures such as reducing crowd travel, closing out the severely affected areas, and providing medical assistance to the severely affected areas can effectively reduce the speed of the outbreak. In addition, the epidemics in Britain and America are still developing rapidly, and isolation measures should continue to be implemented.

References

1. Zhou, T., Liu, Q., Yang, Z., et al.: Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. J. Evid. Based Med. **13**(1), 3–7 (2020). https://doi.org/10.1111/jebm.12376

2. Qianqian, S., Han, Z., Liqun, F., et al.: Study on assessing early epidemiological parameters of coronavirus disease epidemic in China. Chin. J. Epid. **41**(04), 461–465 (2020). https://doi.org/10.3760/cma.j.cn112338-20200205-00069

3. D'Arienzo, M., Coniglio, A.: Assessment of the SARS-CoV-2 basic reproduction number, R0, based on the early phase of COVID-19 outbreak in Italy. Biosaf. Health (2020). https://doi.org/10.1016/j.bsheal.2020.03.004

4. Barlow, N.S., Weinstein, S.J.: Accurate closed-form solution of the SIR epidemic model. Phys. D Nonlinear Phenom. **408**, 132540 (2020). https://doi.org/10.1016/j.physd.2020.132540

5. Guo, C., Guo, W., Chen, C.-H., Wang, X., Liu, G.: The air quality prediction based on a convolutional LSTM network. In: Ni, W., Wang, X., Song, W., Li, Y. (eds.) WISA 2019. LNCS, vol. 11817, pp. 98–109. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30952-7_12

6. Zheng, H., Shi, D.: Using a LSTM-RNN based deep learning framework for ICU mortality prediction. In: Meng, X., et al. (eds.) WISA 2018. LNCS, vol. 11242, pp. 60–67. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02934-0_6

7. Zhuang, Z., Zhao, S., Lin, Q., et al.: Preliminary estimating the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020. Int. J. Infect. Dis. (2020). https://doi.org/10.1016/j.ijid.2020.04.044

8. Zheng, Z., Wu, K., Yao, Z., et al.: The prediction for development of COVID-19 in global major epidemic areas through empirical trends in China by utilizing state transition matrix model (3/10/2020). Available at SSRN: https://ssrn.com/abstract=3552835 or https://doi.org/10.2139/ssrn.3552835

9. Tang, B., Bragazzi, N.L., Li, Q., et al.: An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Infect. Dis. Model. **5**, 248–255 (2020). https://doi.org/10.1016/j.idm.2020.02.001

10. Li, Q., Guan, X., Wu, P., et al.: Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New Engl. J. Med. (2020). https://doi.org/10.1056/NEJMoa2001316