

PAPER • OPEN ACCESS

COVID-19 Propagation Prediction Model Based on Machine Learning

To cite this article: Yunxiang Liu and Yan Xiao 2021 *J. Phys.: Conf. Ser.* **1955** 012094

View the [article online](#) for updates and enhancements.

You may also like

- [ASTEROSEISMIC-BASED ESTIMATION OF THE SURFACE GRAVITY FOR THE LAMOST GIANT STARS](#)

Chao Liu, Min Fang, Yue Wu et al.

- [Noninvasive reconstruction of cardiac transmembrane potentials using a kernelized extreme learning method](#)

Mingfeng Jiang, Heng Zhang, Lingyan Zhu et al.

- [Reservoir rock permeability prediction using support vector regression in an Iranian oil field](#)

Sadegh Saffarzadeh and Seyed Reza Shadizadeh



The Electrochemical Society
Advancing solid state & electrochemical science & technology

243rd Meeting with SOFC-XVIII

Boston, MA • May 28 – June 2, 2023

Accelerate scientific discovery!

Learn More & Register



COVID-19 Propagation Prediction Model Based on Machine Learning

Yunxiang Liu¹, Yan Xiao²

¹Department, Shanghai Institute of Technology, Shanghai, Shanghai, 201418, China

²Department, Shanghai Institute of Technology, Shanghai, Shanghai, 201418, China

*Corresponding author's e-mail: 571128791@qq.com

Abstract. The continuous spread of COVID-19 virus around the world has exerted a great impact on the world's economy and security. China took the lead in responding to the epidemic with a positive attitude and contributed valuable experience to the fight against the epidemic for people around the world. Based on the COVID-19 epidemic data published by the National Health Commission, this study established a time series of the number of newly diagnosed patients on a daily basis. The SIR model of transmission dynamics was used to determine the trend and inflection point of the epidemic, and the SVR algorithm was used to reasonably predict and analyze the number of newly diagnosed patients on a daily basis.

1. Introduction

The continuous spread of COVID-19 virus around the world has exerted a great impact on the world's economy and security. China took the lead in responding to the epidemic with a positive attitude and contributed valuable experience to the fight against the epidemic for people around the world. Based on the COVID-19 epidemic data published by the National Health Commission, this study established a time series of the number of newly diagnosed patients on a daily basis. The SIR model of transmission dynamics was used to determine the trend and inflection point of the epidemic, and the SVR algorithm was used to reasonably predict and analyze the number of newly diagnosed patients on a daily basis.

Quarantine and mass testing policies rapidly reduce the risk of the spread of the outbreak. In this experiment, the epidemic situation in Xinjiang in July 2020 was taken as the data, and the SIR model was used to first determine the epidemic situation and inflection point. Finally, the SVR algorithm was used to select the appropriate interval for training and prediction.

2. Ease of Use

2.1. SIR model

SIR model is a classic transmission model. Where S represents the susceptible person, I represents the infected person and R represents the removed person. Under the SIR model, the population is divided into three categories: the number of susceptible persons is denoted as $S(t)$, which represents the number of people who are not infected but are likely to be infected at time t ; The number of infected persons is denoted as $I(t)$. In this study, it is the confirmed number. The number of displaced persons is denoted as $R(t)$, which represents the number of persons who have been removed from the infected category at time t , including the quarantined persons or the persons who are immune due to cure,



which is simplified as the number of cured persons in this study. According to the theory, the SIR model is established on the basis of the following three basic assumptions: (1) the population always maintains a constant N , that is, the influencing factors such as birth, death and population flow are not considered; (2) In unit time, the number of susceptible persons who can be infected by a patient is directly proportional to the total number of susceptible persons at this time $S(t)$, and the proportional coefficient is β . Thus, it can be concluded that the total number of infected persons in unit time at t is $\beta S(t) I(t)$. (3) The number of patients moved out in unit time is proportional to the number of patients. Assuming the proportional coefficient is γ , the number of patients moved out can be obtained as $\gamma I(t)$. In this study, because only those who were removed included those who recovered, γ could also be called the recovery rate coefficient[4].

The SIR model can be written into ordinary differential equations as shown in Equation (1).

$$\begin{cases} \frac{ds(t)}{dt} = -\beta i(t)s(t) \\ \frac{di(t)}{dt} = \beta i(t)s(t) - \gamma i(t) \\ \frac{dr(t)}{dt} = \gamma i(t) \end{cases} \quad (1)$$

2.2. SVR algorithm

SVR algorithm is to use SVM algorithm to fit the curve for regression analysis and prediction. Classification is the output of discrete values, and regression is the output of a continuous value. SVR algorithm is a tolerant regression model, because the data in the epidemic may be less than the real value due to missed detection and other reasons. If we use a strict linear regression model, it may reduce the accuracy of our prediction.

2.3. Traditional propagation prediction model

Since the emergence of COVID-19, scholars at home and abroad have established a series of prediction models for the epidemic, and achieved certain results. At present, the more popular prediction models mainly focus on dynamics and artificial intelligence models. Dynamic models of COVID-19 propagation mainly include SIR model [5-7].

3. Experiments and Models

3.1. Data processing

The data obtained from the National Health Commission were collected and collated to establish a time series of the number of new people on a daily basis, with the occurrence time as the X axis and the number of new people on a daily basis as the Y axis. The data is shown in the figure 1.

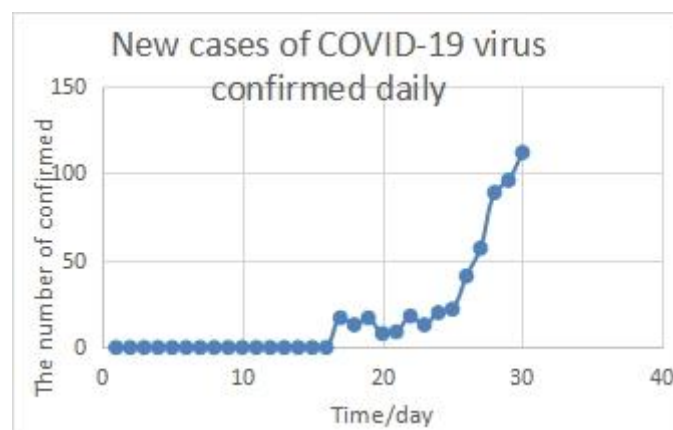


Figure 1. Xinjiang is seeing more confirmed cases every day

3.2. SIR model to predict the trend and inflection point of the epidemic

If the trend and inflection point of the epidemic cannot be judged more accurately, the simple data fitting is meaningless. As time goes by, the epidemic situation will constantly change, and the parameters we set at the initial stage will change. Therefore, we must adjust the parameters in time, fit them in sections, and make dynamic prediction, so as to ensure the accuracy and scientific nature of our prediction. We predetermined the initial parameters based on local conditions and epidemic prevention and control measures. The SIR model was used to solve the epidemic trend and inflection point according to the differential equation.

3.3. The prediction

According to our solution, the outbreak of a turning point selection of training data, and to predict the number of days to choose, in this experiment, for example, we can preliminary judgement, the inflection point of Xinjiang outbreak will appear on August 1, 2020, as a result, we used the data fitting, on July 29 and 30, the number of new diagnosed predicts. Input the data into the SVR algorithm for prediction. The results were compared with those obtained by the simple curve fitting method. The prediction effect is shown in the figure 2.

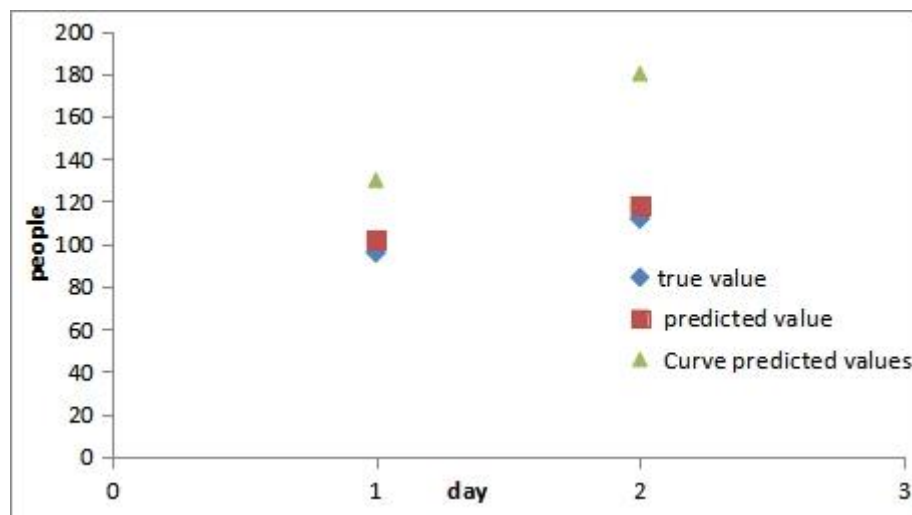


Figure. 2. Prediction renderings

4. Conclusion

The experimental results show that the model combined with SIR and SVR algorithm has higher accuracy and better prediction effect than the simple dynamics and artificial intelligence model. With the increase of China's prevention and control experience and the improvement of its prevention and control mechanism, various powerful measures have quickly cut off the spread of the epidemic, and the model used in this experiment has become more valuable. At the same time, it also shows that the quarantine and other policies currently used by the Chinese government are very effective in COVID-19 prevention and control. As research goes on, data on outbreaks will become more accurate and predictions will become more accurate. The research model will be further optimized.

References

- [1] Natsuko Imai, Anne Cori, Ilaria Dorigatti, et al. Transmissibility of 2019-nCoV[EB/OL]. Imperial College London (25-01-2020), doi: <https://doi.org/10.25561/77148>.
- [2] Effective control measures under the novel Coronavirus fashion trend simulation [J]. Bai Ruhai, Dong Wan Yue, SHI Ying, FENG Aozi, LI Li, XU Ding, LU Jun. Medical knowledge. 2020(02)

- [3] Zhejiang Provincial Health Commission. On February 17, 2020, zhejiang COVID - 19 epidemic situation [EB/OL]. (2020-02-17) [2020-02-18]. https://www.zjwjw.gov.cn/art/2020/2/17/art_1202101_41924242.html.
- [4] Wu Wentao, Li Daning, Li Li, et al. Based on the SIR model, the control measures of different intensities were analyzed A novel coronavirus(2019-nCoV) infection pneumonia epidemic in Wuhan, China To use. *New Info in Medicine*, 2020, 1 (30) : 78-82.
- [5] Zareie B, Mohammad AR, Mansournia A, et al. A model for the development of A new model Prediction of Prediction Parameters Based on China Parameters [J]. *Archives Of Iranian Medicine*, 202, 23 (4): 244-248..
- [6] Sheng Huaxiong, Wu Lin, Xiao Changliang. Analysis and prediction of COVID-19 epidemic transmission model [J] *Journal of Numerical Simulation*, 2020, 32 (5) : 759-766.
- [7] Wang Jianmei, Li Gang. Analysis of the model of COVID-19 non-uniform infectious transmission and intervention [J] *Journal of China University of Science and Technology*, 2020, 49 (3) : 392-398.