Forecasting the long-term trend of COVID-19 epidemic using a dynamic model

Jichao Sun (jichaosun@tencent.com)
Xi Chen
Ziheng Zhang
Shengzhang Lai
Bo Zhao
Hualuo Liu
Ruihui Zhao
Alexander Ng
Yefeng Zheng

Method Article

Keywords: Coronavirus disease 2019, Forecasting model, machine learning, Dynamic-Susceptible-Exposed-Infective-Quarantined

Posted Date: June 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.pex-956/v1

Abstract

The current outbreak of coronavirus disease 2019 (COVID-19) has recently been declared as a pandemic and spread over 200 countries and territories. Forecasting the long-term trend of the COVID-19 epidemic can help health authorities determine the transmission characteristics of the virus and take appropriate prevention and control strategies beforehand. This protocol introduced a new model named Dynamic-Susceptible-Exposed-Infective-Quarantined (D-SEIQ), by making appropriate modifications of the Susceptible-Exposed-Infective-Recovered (SEIR) model and integrating machine learning based parameter optimization under epidemiological rational constraints. The model could be applied to forecast the long-term trend of the current COVID-19 epidemic. The parameters obtained by the model could help assess the effectiveness of prevention and intervention strategies on epidemic control in different countries.

Introduction

Coronavirus disease 2019 (COVID-19) is an infectious pneumonia caused by severe acute respiratory syndrome coronavirus 2. The disease was first reported in December 2019 in Wuhan city, the capital of Hubei province in China, and has since then spread across China and globally. As of 6 April 2020, a total of 1.27 million COVID-19 cases and 69,400 deaths have been reported in more than 200 countries and territories. The World Health Organization (WHO) has declared the COVID-19 outbreak as a Public Health Emergency of International Concern and a pandemic recently.

Forecasting long-term trend of the epidemic can help health authorities determine the transmission characteristics of the virus and take appropriate prevention and control strategies beforehand. Recently, some researchers applied the traditional epidemic models like Susceptible-Exposed-Infective-Recovery (SEIR) or machine learning models like logistic regression to fit the trend of COVID-19. To the best of our knowledge, most of those researches were performed retrospectively, or subject to overfitting or under-fitting problems. The validity of SEIR model depends on accurate estimation of virus transmission characteristics such as the basic reproduction number R_0, incubation period and infectious period. In a real scenario, those parameters are not easy to estimate. On the other hand, due to scarcity of training data and valid features, machine learning models were subject to overfitting, restricted to retrospective analysis, or only forecasting short-term trends.

To address the aforementioned issues, we propose a novel model named Dynamic-Susceptible-Exposed-Infective-Quarantined (D-SEIQ), by making appropriate transformations of the SEIR model and integrating machine learning based parameter optimization under reasonable constraints, which improved the performance of long-term trend forecast for COVID-19 in China. In addition, the model parameters could provide insights into the analysis of COVID-19 transmission characteristics and the effectiveness of interventions.

Reagents

Not applicable because this is a computational modeling study

Equipment

Not applicable because this is a computational modeling study

Procedure

D-SEIQ model

The primary differences from our D-SEIQ model and SEIR model are 1) replacing recovered individuals R with quarantined individuals Q, and 2) introducing dynamics to the calculation of the effective reproduction number R_t that is dependent on time.

Some previous work employed the traditional SEIR model, which assumes that the exposed individuals (who have been infected but no symptoms yet) are not infectious. However, it is reported that COVID-19 might be transmissible for exposed individuals . Besides, due to lack of specialized treatment, the infectious period should not be interpreted as time between infective (I) and recovered (R) but time between infection (I) and quarantined (Q). Therefore, we proposed to replace the recovered individuals R with the quarantined individuals Q and the model became the SEIQ model. The quarantined individuals Q indicated the confirmed cases who were centrally quarantined. The epidemic spreading model for the SEIQ model is shown in Figure 1.

The transmission dynamics are governed by the following system of equations:

Eq(1) (Equation can not be displayed here, please refer to supplementary file)

where $N = S(t) + E(t) + I(t) + Q(t)$ is the total population, which is assumed consistent.

Like the SEIR model, parameter β indicates the infectious rate with $\beta = Rt/TE$ where R_t is the dynamic effective reproduction number and TE is the average duration of incubation; parameter σ indicates the incubation rate with $\sigma = 1/TE$. However, in our model, parameter γ indicates the quarantine rate with $\gamma = 1/TI$ (where TI is average duration of an infectious individual to be detected and quarantined). The parameter TI may vary across different regions and the difference reflects the timeliness of patient detection and admission.

The basic reproduction number R_0 is the most important parameter to determine the intrinsic transmissibility of COVID-19, and it is defined as the average number of infections one infectious agent can generate over the course of the infectious period without any interventions. R_0 was assumed to be a constant or arbitrarily modified at specific points for forecasting in previous work [12, 13]. However, in real-world scenarios, with the development of epidemic, more and more interventions are often taken to control the spread, which gradually reduce R_0. In this work, the basic reproduction number R_0 is generalized to a dynamic value R_t, which is defined as the average number of secondary infectious cases generated by an infectious at time t. After the worldwide outbreak of COVID-19, many governments took considerable measures to contain the spread of the virus. In our preliminary analysis and similar to previous work [14], the infectious rate β was shown to decrease exponentially with time. As parameter TE is constant, the effective reproduction number R_t should follow similar pattern as decreasing exponentially with time. Thus, we introduced time-dependent dynamics to the calculation of R_t for better simulation of the real-world transmission,

Eq (2) (Equation can not be displayed here, please refer to supplementary file)

where R∞ is the final reproduction number at the end of the pandemic and θ is the decrease ratio of the reproduction number, which is associated with the corresponding interventions. When t = 0, R_t=R_0, and it gradually reduces to R∞. The epidemic is considered to be under control with R_t < 1, and the reasonable range of R∞ was referred to some previous analysis of coronavirus.

Parameter constraints and optimization

The simulation and prediction of the D-SEIQ model requires determination of the parameters R_0, R∞，TE, TI, θ. Although we incorporated machine learning to help us to fit the reported data, the parameter range needs to be set carefully and to conform to epidemiological rationality. For instance, Wu et al. applied an adjusted SEIR model to estimate R_0 (R_0= 2·68) in major cities of China by analyzing the number of cases exported from Wuhan internationally. Some work concluded that the daily reproduction number varied between 2 and 7. Therefore, we set the reasonable range for parameter R_0 to be [2, 7]. Likewise, after reviewing the previous work on the analysis of COVID-19, we summarized the ranges for parameters in our model as Table 1 (supplementary file). And, we set TE > TI as additional constraints. Therefore, the parameter optimization process is as follows:

a. Initialize the number of confirmed cases Q at time t = 0 according to the official report.

b. Initialize the parameters R_0, R∞, TE, TI, θ

c. Calculate the time-dependent effective reproduction number R_t

d. Solve ordinary differential equations in Equation (1) to determine E(t), I(t), Q(t)

e. Set loss function as the sum of mean squared errors of daily and cumulative confirmed numbers, and then estimate the parameters R_0, TE, TI, R∞, θ based on grid search with dynamically adapted search steps to obtain the best D-SEIQ model at time t.

Data processing

We obtained the updated data of the cumulative confirmed cases from the National Health Commission (NHC) of the People's Republic of China. The newly confirmed cases were also collected on a daily basis. Considering that medical resources and interventions might vary in different regions, we fitted our model on the data from three different regions: 1) China excluding Hubei, 2) Hubei excluding Wuhan, and 3) Wuhan.

Moreover, we adjusted the number of newly confirmed cases in Wuhan between 12 February and 14 February, due to the inclusion of clinically confirmed cases without coronavirus test. The clinically confirmed cases between 12 February and 14 February were assumed to be suspicious cases in last 7 days. Specifically, we redistributed the clinically confirmed cases according to the distribution of suspected cases over the past 7 days.

Troubleshooting

We identified two issues that may arise during the implementation of the protocol: data processing issue and parameter learning issue.

1) Data processing:

In China, the criteria for ascertainment of COVID-19 was changed during Feb 12 to Feb 14,

due to the inclusion of clinically confirmed cases without coronavirus test.

We arbitrarily redistributed the clinically confirmed cases according to the distribution of suspected cases over the past 7 days. However, one can redistributed the numbers using different approaches, only if the process is sound and reasonable. Moreover, when applying the model to other countries, one should pre-process the data accordingly to make the reported number less biased.

2) Parameter learning:

We set the reasonable range for parameters using the current available evidence by April.

If any evidence changed through time, one should adjust the reasonable range accordingly if it is scientific valid.

Time Taken

Not applicable

Anticipated Results

With the proposed D-SEIQ model and accumulating COVID-19 cases in different country, one can forecast the long-term trend of COVID-19 and expected turning point, which might serve as a reflection of containment measures currently taken in different areas. Moreover, the learned parameters provides estimates of virus transmission characteristics, including basic reproduction number, incubation period and infection period. Those parameters help authorities determine the transmission potential of the new virus and make appropriate intervention strategies.

References

1 Zhou X, Hong N, Ma Y, et al. Forecasting the Worldwide Spread of COVID-19 based on Logistic Model and SEIR Model. *medRxiv* 2020: 2020.03.26.20044289.

2 Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 2020; **395**(10225): 689-97.

3 Tátrai D, Várallyay Z. COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability. *arXiv e-prints*, 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200314160T (accessed March 01, 2020).

4 Roosa K, Lee Y, Luo R, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling* 2020; **5**: 256-63.

Acknowledgements

Figures

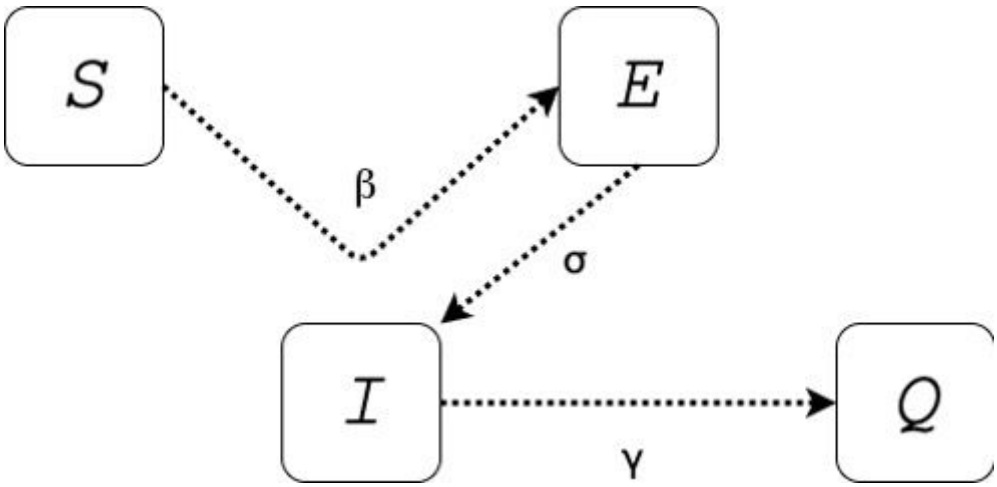

Figure 1

Figure 1 Epidemic spreading diagram for SEIQ model S: susceptible; E: exposed; I: infective; Q: quarantined; β indicates the infectious rate. Parameter σ indicates the incubation rate with σ = 1/TE (incubation period). Parameter γ indicates the quarantine rate with γ = 1/TI (infectious period).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- protocol.docx