

# 基于新冠肺炎预测模型选择及防疫能力评价的研究与分析

陈林, 方百青, 潘纯权, 肖俊, 康茜

重庆科技学院, 数理与大数据学院, 重庆

收稿日期: 2023年5月28日; 录用日期: 2023年6月23日; 发布日期: 2023年6月30日

## 摘要

随着2020年新冠肺炎疫情在全球范围内的爆发, 全球面临了前所未有的挑战。本文旨在研究各国采取的防疫措施的有效性, 并探讨有效的防疫措施对疫情控制、经济复苏和人民生命安全的重要性。本文基于新冠肺炎预测模型选择及防疫能力评价, 采用多种分析方法进行研究和分析。首先, 本文采用时间序列ARIMA模型对疫情进行分析预测, 并利用经典的传染病动力学模型, 加入潜伏患者和死亡概率, 建立SEIRD模型对疫情进行预测。接着, 采用多种机器学习模型与滚动预测方法相结合, 提升模型的精确度, 并通过平均绝对误差(MAE)度量模型的预测精度。然后, 利用多变量LSTM对疫情进行预测, 并基于LSTM深度学习和多种机器学习的基础上建立赋权组合模型的方法, 得到更为精准的预测模型。最后, 本文基于多元线性回归模型分析感染率的影响因素, 考虑人口密度分布的情况下, 量化评价我国各个区域的防疫能力, 并从不同角度提出疫情防控工作的可行性建议, 形成精细的防疫模式。本文研究结果表明, 有效的防疫措施可以有效控制疫情的传播, 减少患者数量和死亡率, 并为经济复苏提供了条件。此外, 基于多种分析方法和模型选择, 本文提出了更为有效的疫情预测模型和防疫能力评价模型, 为应对未来的传染病疫情提供了重要的参考依据。

## 关键词

新冠肺炎, 预测模型, 防疫能力评价, ARIMA, 机器学习

# Research and Analysis on the Selection of COVID-19 Prediction Models and Evaluation of Epidemic Prevention Capability

Lin Chen, Baiqing Fang, Chunquan Pan, Jun Xiao, Qian Kang

Mathematics and Big Data College, Chongqing University of Science and Technology, Chongqing

Received: May 28<sup>th</sup>, 2023; accepted: Jun. 23<sup>rd</sup>, 2023; published: Jun. 30<sup>th</sup>, 2023

## Abstract

The outbreak of COVID-19 in 2020 has posed unprecedented challenges to the world on a global scale. This article aims to investigate the effectiveness of epidemic prevention measures taken by different countries, and explore the importance of effective epidemic prevention measures in controlling the epidemic, promoting economic recovery, and ensuring people's safety. This article conducts research and analysis through various analytical methods based on the selection of COVID-19 forecasting models and the evaluation of epidemic prevention capabilities. Firstly, this article utilizes a time series ARIMA model to analyze and predict the epidemic, and incorporates latent patients and mortality rates into the classical infectious disease dynamics model to establish the SEIRD model for epidemic forecasting. Furthermore, a combination of various machine learning models and rolling forecasting methods is used to improve the accuracy of the model, and the Mean Absolute Error (MAE) is used to measure the predictive accuracy of the model. Subsequently, the epidemic is predicted using multivariate LSTM, and a weighted combination model is established based on LSTM deep learning and various machine learning methods to obtain a more accurate predictive model. Finally, this paper analyzes the influencing factors of infection rate based on the multiple linear regression model, quantitatively evaluates the epidemic prevention capabilities of various regions in China considering population density distribution, and proposes feasible suggestions for epidemic prevention and control from different perspectives, forming a refined epidemic prevention mode. The research results of this paper indicate that effective epidemic prevention measures can effectively control the spread of the epidemic, reduce the number of patients and mortality, and provide conditions for economic recovery. In addition, based on multiple analysis methods and model selection, this paper proposes more effective epidemic prediction models and epidemic prevention capacity evaluation models, providing important reference for coping with future infectious disease epidemics.

## Keywords

COVID-19, Predictive Model, Evaluation of Epidemic Prevention Capabilities, ARIMA, Machine Learning

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

COVID-19 疫情自 2020 年全球爆发以来[1], 对全球经济和人民的生命安全造成了极大的威胁。随着疫情不断发展, 各国政府采取了各种防疫措施来应对疫情, 然而这些措施的有效性并不相同。因此, 本文通过新冠肺炎预测模型选择和防疫能力评价, 对各国防疫措施的有效性进行了研究和分析。

本文采用时间序列 ARIMA 模型对疫情进行了分析预测[2]。在经典的传染病动力学模型下, 本文结合潜伏患者和死亡率等因素, 建立了 SEIRD 模型对疫情进行预测。此外, 本文还使用多种机器学习模型与滚动预测方法相结合, 从而提高模型的精度, 并通过平均绝对误差(MAE)来度量模型的预测精度。最后, 本文还采用多变量 LSTM 模型对疫情进行预测, 并在 LSTM 深度学习和多种机器学习的基础上建立了赋权组合模型, 以得到更加准确的预测结果。本文还分析了感染率的影响因素, 并采用多

元线性回归模型分析各区域的防疫能力。在考虑人口密度分布的情况下，本文量化评价了我国各个区域的防疫能力，并从感染率的影响因素的不同角度提出了疫情防控工作的可行性建议，形成了一种精细的防疫模式。结果表明，有效的防疫措施可以有效控制疫情的传播，减少患者数量和死亡率，并为经济复苏提供了条件。在本文所采用的预测模型中，赋权组合模型的预测效果最好。因此，我们可以得出结论，通过综合运用时间序列分析、传染病动力学模型、机器学习算法和多元线性回归模型等方法，可以有效预测疫情，并评估各国防疫措施的有效性。另外，本研究还分析了各国在疫情期间采取的不同防疫措施的有效性，并结合 SEIRD 模型的预测结果进行评估。结果显示，一些措施如限制人员流动、加强社交距离等能够有效控制疫情传播，而一些措施如禁止公共聚集、关闭商业场所等虽然能够减少疫情传播，但对于经济影响较大。此外，一些措施如检测、隔离和追踪确诊病例等也被证明是有效的防疫手段。

综合分析结果，本研究认为，有效的防疫措施可以有效控制疫情的传播，减少患者数量和死亡率，并为经济复苏提供了条件。因此，各国应当继续加强防疫措施，包括但不限于限制人员流动、加强社交距离、加强检测和隔离等，以确保疫情得到控制，并在逐步恢复经济的过程中采取有效的措施，以避免出现二次爆发。另外，本研究提出了一些针对感染率影响因素的建议，如加强公共卫生意识，提高人们的健康素养等。总之，本研究通过多种预测模型和评价方法，对新冠肺炎疫情进行了全面研究和分析，为各国在防疫措施上提供了有益的参考，对于应对突发公共卫生事件具有一定的指导意义。同时，本研究还发现了一些有价值的问题，如各种防疫措施的有效性和经济影响等，这些问题值得进一步深入研究。

## 2. 预测方法

我们采用 SERID、机器学习和 LSTM 分别预测疫情，根据得到的结果综合分析，选择一个最好的模型来预测疫情。

### 2.1. SEIRD 预测

在已有的新冠疫情预测模型成果基础上，我们通过大量搜集和阅读本课题相关资料和数据，针对不同模型进行了相关的优化，主要对四个方面进行了研究：

#### 2.1.1. 传统 SEIR 模型

SEIR 是在 SIR 模型上面进行改进。SIR 是易感者  $S$  被传染者  $I$  所感染，然后被治愈好称为移除者  $R$  (即治愈好不会在被感染的和死亡的)。在此基础上，如果所研究的传染病有一定的潜伏期，与病人接触过的健康人并不马上患病，而是成为病原体的携带者，归入  $E$  类[3]。此时可以建立微分方程：

$$\frac{dS}{dt} = -\frac{r\beta SI}{N} \quad (1)$$

$$\frac{dE}{dt} = \frac{r\beta SI}{N} - \sigma E \quad (2)$$

$$\frac{dI}{dt} = \sigma E - \gamma I \quad (3)$$

$$\frac{dR}{dt} = \gamma I \quad (4)$$

守恒关系  $S(t) + E(t) + I(t) + R(t) =$  人口总数。

根据查阅资料设置初始参数，绘制出以下的 SEIR 图，见图 1：

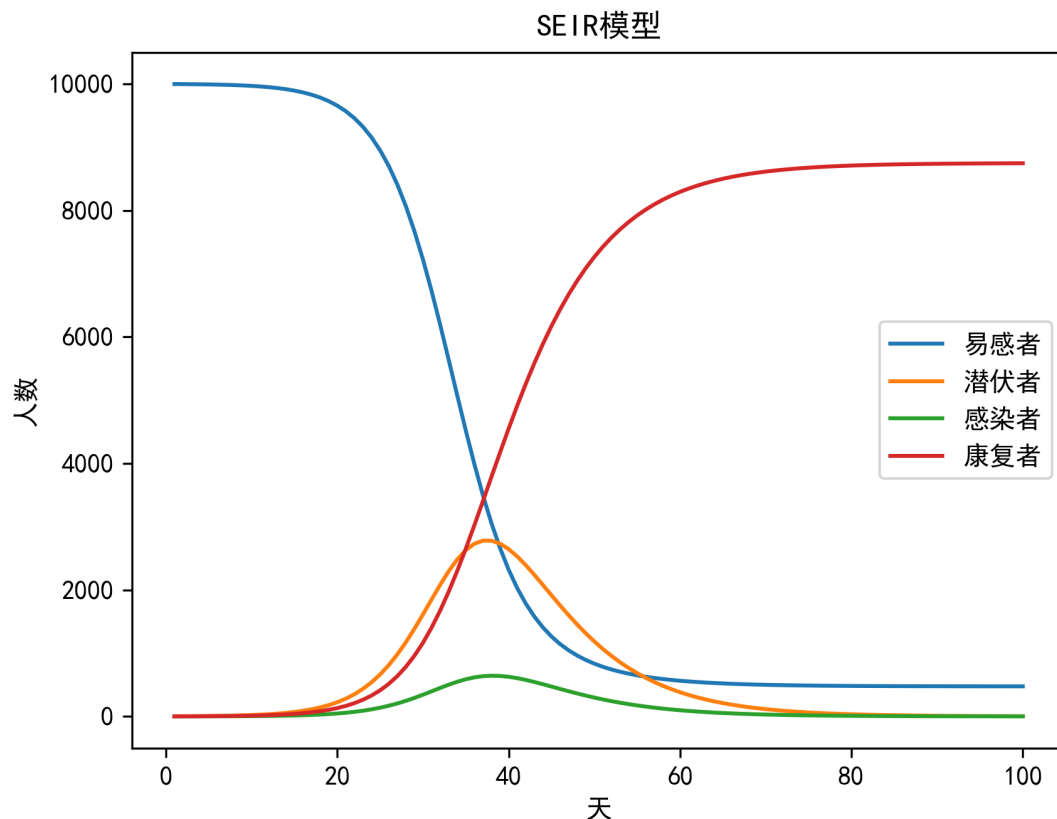


Figure 1. The conventional SEIR model  
图 1. 传统 SEIR 模型

### 2.1.2. 基于改进后的 SEIRD 模型

根据疫情期间的报告,存在着很多潜伏患者传染人的事件,所以传统的 SEIR 解决不了这类问题,所以我们基于 SEIR 上加上潜伏期患者感染易感者的概率  $\beta_2$  和每天患者接触的人  $r_2$ , 所以得到每天潜伏期患者传染  $\frac{r_2\beta_2ES}{N}$  人。在此之外,在疫情严重的时候,死亡人数也很多,为了方便研究,在 SEIR 上加上了死亡概率  $p$ , 得到每天死亡人数  $pI$  [4] [5]。

以下为符号说明, 以及根据查询的资料设置参数, 然后确定概率参数、设计 SEIRD 模型。

建立微分方程:

$$\frac{dS}{dt} = -\frac{r\beta SI}{N} - \frac{r_2\beta_2 ES}{N} \quad (1)$$

$$\frac{dE}{dt} = \frac{r\beta SI}{N} - \sigma E + \frac{r_2\beta_2 ES}{N} - \gamma_2 E \quad (2)$$

$$\frac{dI}{dt} = \sigma E - \gamma I - pI \quad (3)$$

$$\frac{dR}{dt} = \gamma_1 I + \gamma_2 E \quad (4)$$

$$\frac{dD}{dt} = pI \quad (5)$$

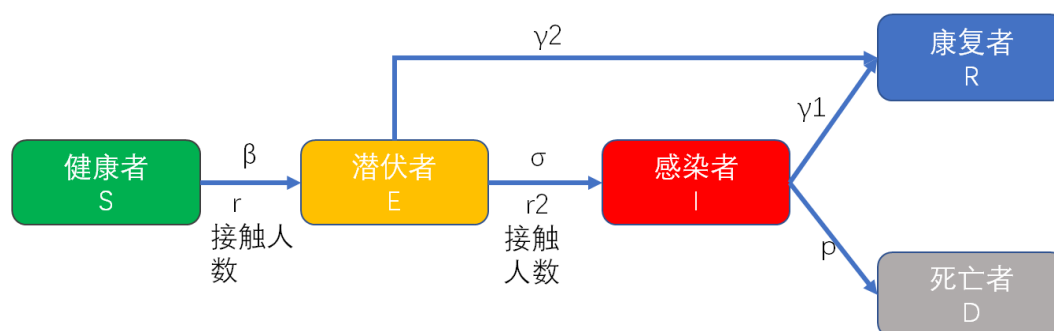
综合分析每个参数, 并根据实际情况, 给出每个符号的初值, 见表 1。

**Table 1.** Symbol notation  
**表 1.** 符号说明

符号	涵义解释	初值设定
$S$	易感者	9999
$E$	潜伏者(病毒存在潜伏期)	0
$I$	感染者	1
$R$	康复者	0
$\beta$	易感者被感染者感染成潜伏者的概率	0.1
$\beta_2$	易感者被潜伏者感染成潜伏者的概率	0.03
$\sigma$	潜伏者变为感染者的概率(有自愈的情况)	0.14
$\gamma_1$	感染者康复的概率	0.81
$\gamma_2$	潜伏者康复的概率	0.05
$r$	一个感染者在一天内遇见的人	10
$r_2$	一个潜伏者在一天内遇见的人	20
$N$	一个地区的总人数	10000
$p$	死亡的概率	0.02

采集印度的 COVID-19 感染者统计数据进行分析,将 COVID-19 感染者统计数据和采用 SEIRD 模型模拟的结果进行对比,从数据走势可看出,实际数据曲线趋势与 SEIRD 曲线趋势相同,表明 SEIRD 模型能够有效反映 COVID-19 感染者数量的发展趋势。COVID-19 存活时间为 2~14 d,将平均潜伏期定为符合报道的 7 d,死亡率设置为 2%。在采取严格的防控措施下,感染者数量峰值显著降低,疫情高峰更早地到来。采用 SEIRD 模型,对重新调优后的模拟感染者曲线进行网格搜索,寻找最优参数。实验得出,  $\beta = 0.1$ 、 $\beta_2 = 0.03$ 、 $\sigma = 0.14$ 、 $\gamma_1 = 0.81$ 、 $\gamma_2 = 0.05$ , 这条曲线拟合效果最好[6] [7]。

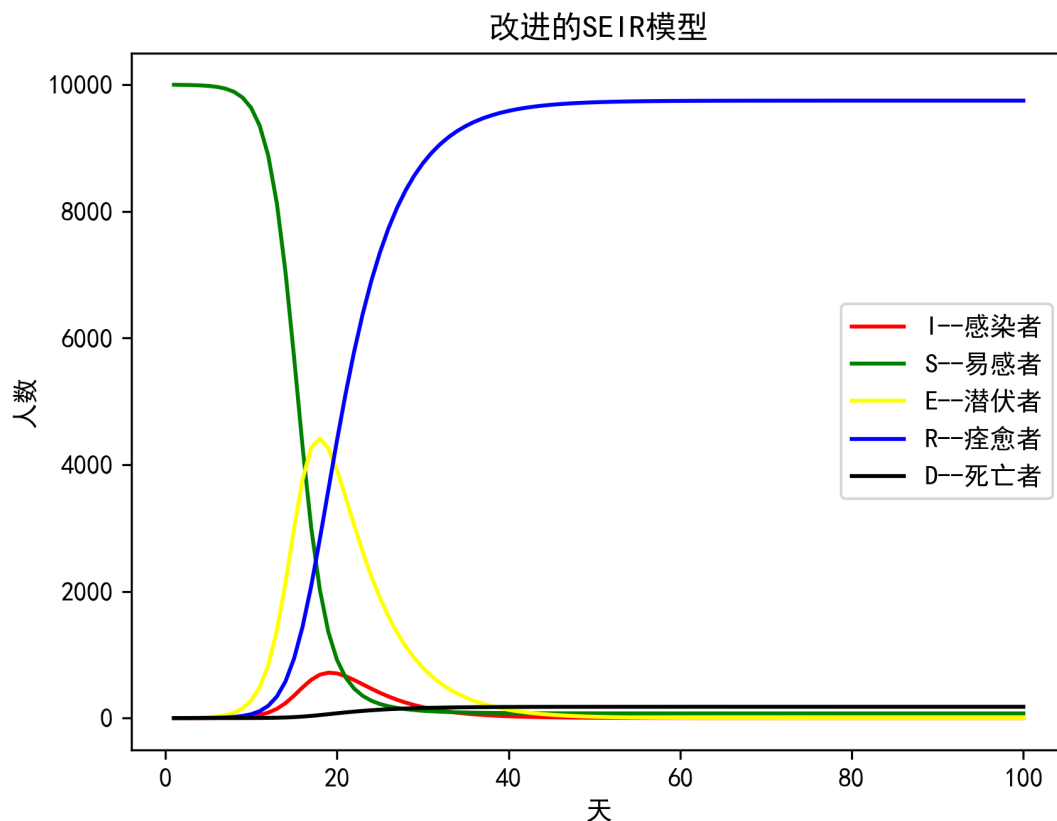
如下,得出模型示意图,如图 2:



**Figure 2.** Diagram of the model  
**图 2.** 模型示意图

则根据以上的参数绘制的 SEIRD 图,见图 3。

分析该模型的特点,在 23 天的时候感染者达到最大值,然后第 50 天的时候疫情已经比较平稳,此时几乎没有或者很少的人被感染。从图上得知,在 21 天的时候潜伏者的数量达到最大值,到感染者达到峰值有 2 天的时间,说明新冠潜伏者可能就是在患病的第二天就开始传染[8] [9]。综合考虑,得出结论:



**Figure 3.** The SEIRD model  
**图 3.** SEIRD 模型

- 1) 控制  $r$  和  $r_2$ , 即控制感染者和潜伏者接触到的人, 所以需要疫情需要将感染者和潜伏者重点隔离。
- 2) 控制  $\beta$  和  $\beta_2$ , 即加强预防疫情, 戴口罩, 多消毒[10] [11]。
- 3) 减小  $\sigma$  和增大  $\gamma_2$ , 即减小潜伏期患者升级为患者的概率和治愈概率。根据研究表明潜伏者增加身体素质能力, 保持乐观积极向上的态度, 能增加自愈的能力。
- 4) 加大  $\gamma_1$ , 即增大患者治愈的概率。政府医院需要加大对患者的治愈能力[12] [13] [14]。

## 2.2. 机器学习预测

作为数据处理和预测的工具, 机器学习算法具备自动学习和提高预测准确率的能力, 无需手动调整参数和特征。它可以快速地处理大规模数据, 对海量数据进行分析 and 处理, 从而提高数据处理的效率和质量。此外, 机器学习算法可以处理非线性关系和复杂的数据结构, 并捕捉数据之间的关系, 从而提高预测的准确性。这种算法可以应用于多个领域, 如金融、医疗、交通等。在这些领域中, 它可以提高预测的准确性和效率, 减少人工干预和错误率, 为决策提供更加科学的依据。因此, 我们选择机器学习算法作为我们的数据处理和预测工具[15]。

### 2.2.1. 数据来源和预处理

本研究使用了来自中国卫生部的官方数据, 包括每日确诊病例数、每日治愈病例数、每日死亡病例数、每日新增疑似病例数等数据。这些数据是从 2020 年 1 月 1 日开始的, 并持续更新到疫情结束。在数据处理方面, 我们首先对数据进行了清洗, 包括去除缺失值、异常值和重复值。然后, 我们对数据进行了可视化分析, 以更好地了解疫情的发展趋势和规律。最后, 我们使用 Python 编程语言将数据进行了特



征工程和标准化处理，以便于机器学习算法的应用。

### 2.2.2. 机器学习算法

本研究使用了三种机器学习算法来预测中国新冠肺炎疫情的结果，分别是线性回归、决策树和支持向量机。

线性回归是一种广泛应用于预测的算法，它建立了一个线性模型来预测输出变量和输入变量之间的关系。在本研究中，我们将每日新增确诊病例数作为输出变量，将每日新增疑似病例数、每日治愈病例数和每日死亡病例数作为输入变量进行建模和预测。

决策树是一种基于树形结构的分类和回归算法，它将数据集分成一些小的子集，并生成树形结构，其中每个叶子节点代表一个类别或一个数值。在本研究中，我们将每日新增确诊病例数作为输出变量，将每日新增疑似病例数、每日治愈病例数和每日死亡病例数作为输入变量进行建模和预测。

支持向量机是一种用于分类和回归的机器学习算法，它利用高维空间中的超平面将数据分成两个类别或预测一个数值。在本研究中，我们将每日新增确诊病例数作为输出变量，将每日新增疑似病例数、每日治愈病例数和每日死亡病例数作为输入变量进行建模和预测。

### 2.2.3. 预测感染人数

我们使用了三种机器学习算法来预测每日新增确诊病例数，并将预测结果与实际数据进行比较。结果表明，支持向量机算法的预测效果最好，其均方根误差为 0.2829，决定系数为 0.87。线性回归算法的预测效果次之，其均方根误差为 0.3159，决定系数为 0.81。决策树算法的预测效果最差，其均方根误差为 0.4756，决定系数为 0.62。

利用支持向量机预测得到的结果，见图 4，图 5。

虽然机器学习的结果能反映出新增人数的大致的趋势，但是预测的误差比较大，说明机器学习还需要再进一步的改进。

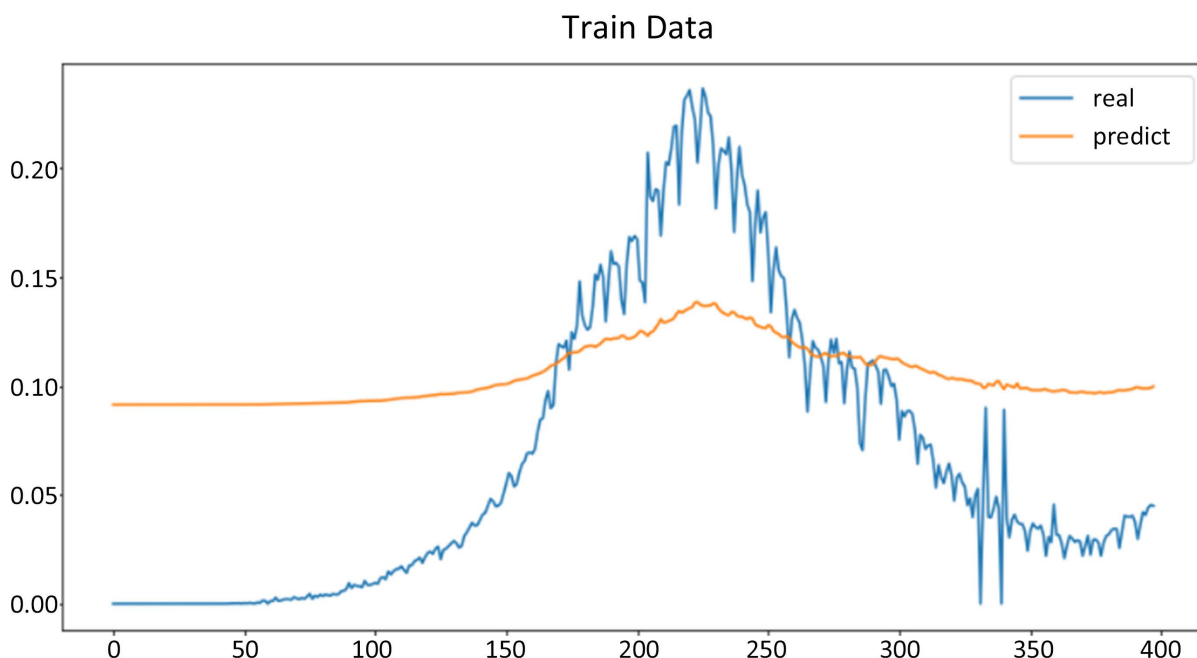


Figure 4. Prediction of support vector machine on the training set

图 4. 支持向量机在训练集上的预测

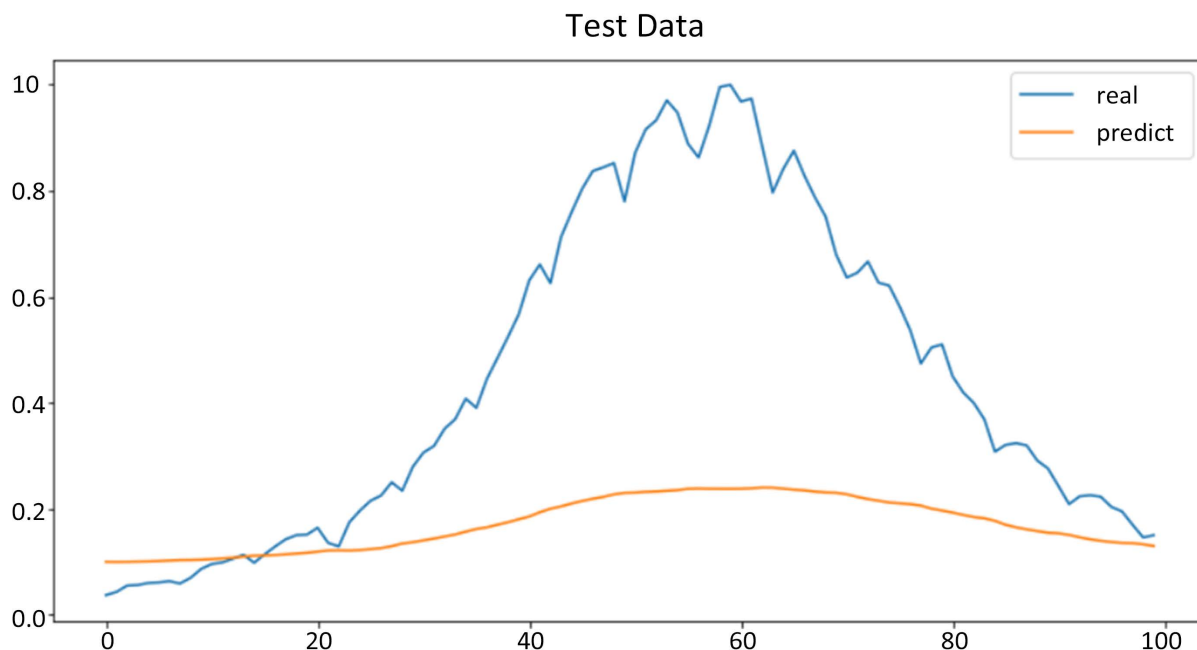


Figure 5. Prediction of support vector machine on the test set

图 5. 支持向量机在测试集上的预测

#### 2.2.4. 结论与建议

本研究使用机器学习算法预测了中国新冠肺炎疫情的发展趋势，并分析了影响疫情发展的因素。我们建议政府加强人口流动管控，优化交通状况，增加医疗资源投入，并制定更加有力的政策措施来应对疫情。

此外，我们也需要继续研究和改进机器学习算法，以提高预测的准确性和可靠性。同时，我们需要更加深入地分析不同地区的疫情发展规律和影响因素，以制定更加精细化、有针对性的防控措施，保障人民身体健康和社会稳定[16] [17] [18]。

### 2.3. LSTM 预测

由于 CNN (卷积神经网络)的特征检测层通过训练数据进行学习，所以在使用 CNN 时，避免了显示的特征抽取，而隐式地从训练数据中进行学习；再者由于同一特征映射面上的神经元权值相同，所以网络可以并行学习，这也是卷积网络相对于神经元彼此相连网络的一大优势。卷积神经网络以其局部权值共享的特殊结构在语音识别和图像处理方面有着独特的优越性，其布局更接近于实际的生物神经网络，权值共享降低了网络的复杂性，特别是多维输入向量的图像可以直接输入网络这一特点避免了特征提取和分类过程中数据重建的复杂度。LSTM 则是特殊的 CNN，其具有按时间顺序扩张的特性，在后文的感染人数预测方面具有更好的优越性。

#### 2.3.1. LSTM 介绍

长短时记忆网络(Long short-term memory, LSTM)是一种循环神经网络(Recurrent neural network, RNN)的特殊变体，具有“门”结构，通过门单元的逻辑控制决定数据是否更新或是选择丢弃，克服了 RNN 权重影响过大、容易产生梯度消失和爆炸的缺点，使网络可以更好、更快地收敛，能够有效提高预测精度。它具有记忆功能，所以适合预测时间序列。由遗忘门，输入门，输出门构成。

首先将数据切分，如图 2~6，将数据切分为  $i$  个样本，时间步长为 14，每一个时间步长表示为  $x_t$ ，即有  $x_1, x_2, \dots, x_{14}$ 。并且每个时间步长为 3，即[传染率，治愈数，死亡数]。总结起来这个数据集的格式变为



(i,14,3), 见图 6。

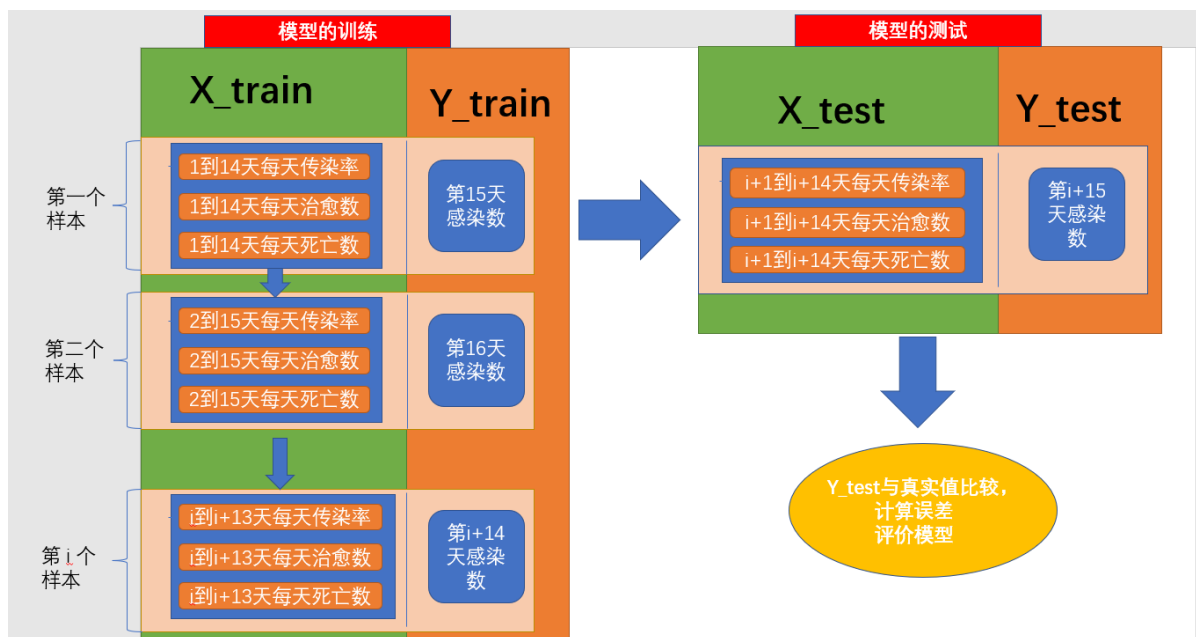


Figure 6. The principle of data splitting  
图 6. 数据切分原理

LSTM 的运行原理, 见图 7。

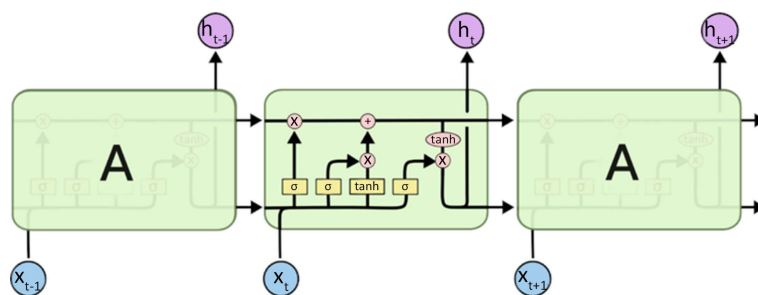


Figure 7. Operating principle of LSTM  
图 7. LSTM 运行原理

a) 遗忘门

遗忘门决定了来自上一时间步的单元状态  $C_{t-1}$  中的多少信息需要忘掉, 见图 8

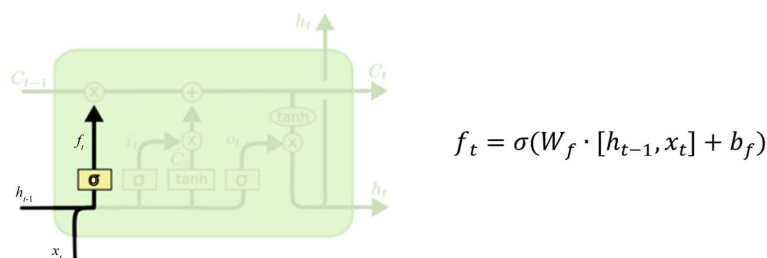


Figure 8. The memory loss gate  
图 8. 遗忘门

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

其中遗忘门输入的是上一状态的隐藏状态，也就是上一时间步的结果  $h_{t-1}$  和当前时间步  $x_t$ 。 $\sigma$ 表示激活函数，作用是将值映射到 0~1 之间。 $W_f$ 是权重， $b_f$ 是偏置。

最后将得到的结果  $f_t$  乘以上一步单元状态  $C_{t-1}$ 。这里的  $f_t$  的范围是在 [0,1]。当  $f_t$  趋近于 0 时，表示忘记所有记忆，同理趋近于 1 时，保留所有记忆。

b) 输入门

输入门的原理和遗忘门一样。输入门决定了当前处理的时间步中多少信息需要加入到当前的单元状态  $C_t$  中。见图 9。

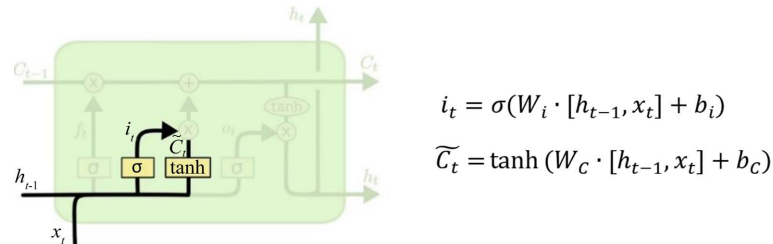


Figure9. Input gate  
图 9. 输入门

根据输入的  $h_{t-1}$  和  $x_t$  得到  $i_t$ ,  $i_t$  的范围也是在 [0,1]。同理可以得到  $\tilde{C}_t$ ，这里的  $\tanh$  也是一个激活函数，将值映射到 [-1,1] 之间。最后，见图 10，将得到的  $\tilde{C}_t$  与输入门的输出值  $i_t$  相乘，并与经过遗忘门处理的来自上一步的单元状态 ( $f_t C_{t-1}$ ) 相加，作为下一个时间步的单元状态  $C_t$ 。

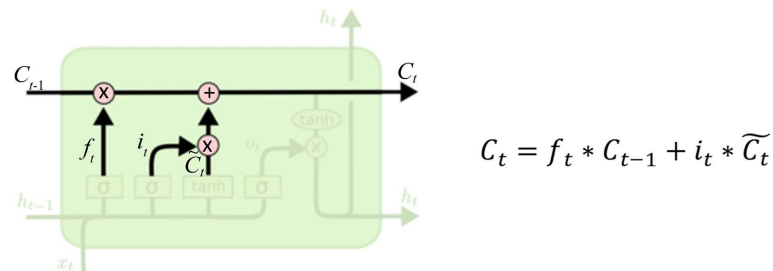


Figure10. Obtain a new cell state  
图 10. 得到新的单元状态

c) 输出门

输出门决定当前单元状态  $C_t$  中的多少信息作为处理当前时间步的输出值  $y_t$  与隐藏状态的值  $h_t$ 。见图 11。

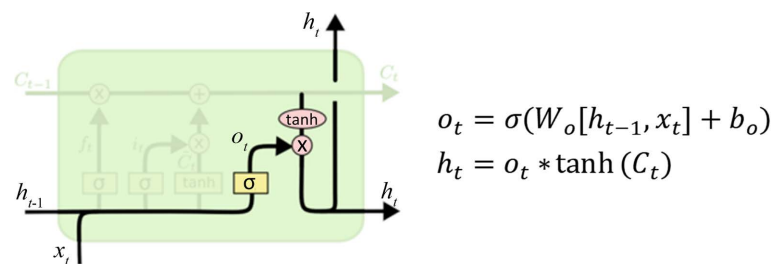


Figure11. Output  $h_t$   
图 11. 输出  $h_t$

### 2.3.2. 数据清洗, 切割数据

结合图 6, 我们选取印度的数据, 假设时间段为 14 天, 即时间步长等于 14, 介绍 LSTM 多变量模型的实际操作步骤:

- 1) 按照训练集: 仿制滚筒模型, 首先选取第 1 天到第 14 天每天传染率、治愈数、死亡数作为第一个混合的训练集, 第 15 天的感染数作为测试集; 第 2 天到第 15 天每天传染率、治愈数、死亡数作为第二个混合的训练集, 第 16 天的感染数作为测试集, 一直怎样往复下去, 直至第  $i$  个训练集。
- 2) 使用训练好的 LSTM 模型对第  $i+1$  天到  $i+14$  天新增确诊进行预测。
- 3) 对预测的结果和真实的结果计算 MSE 误差(均方误差)。
- 4) 重复以上步骤, 选取不同的时间步长, 计算出误差, 最后找到最小的误差, 最为最终的时间步长。

### 2.3.3. 调参

调整参数时, 采用 GridSearchCV 网格搜索寻找 LSTM 模型最优参数, 以及手动调节训练的步长。

1) `batch_size` 的选择。`batch_size` 定义为一次训练所选取的样本数。`batch_size` 的大小影响模型的优化程度和速度。同时其直接影响到 GPU 内存的使用情况, 假如你 GPU 内存不大, 该数值最好设置小一点。对 `batch_size` 在 5 到 32 中进行网格搜索, 得到最好的 `batch_size` 为 6 最好。

2) 迭代次数的选择。首先假定迭代次数为 300。按照上述样本和 `batch_size` 等于 6 进行训练, 得到的训练集的误差和测试集的误差, 见图 12。

关于预测indie的损失折线图

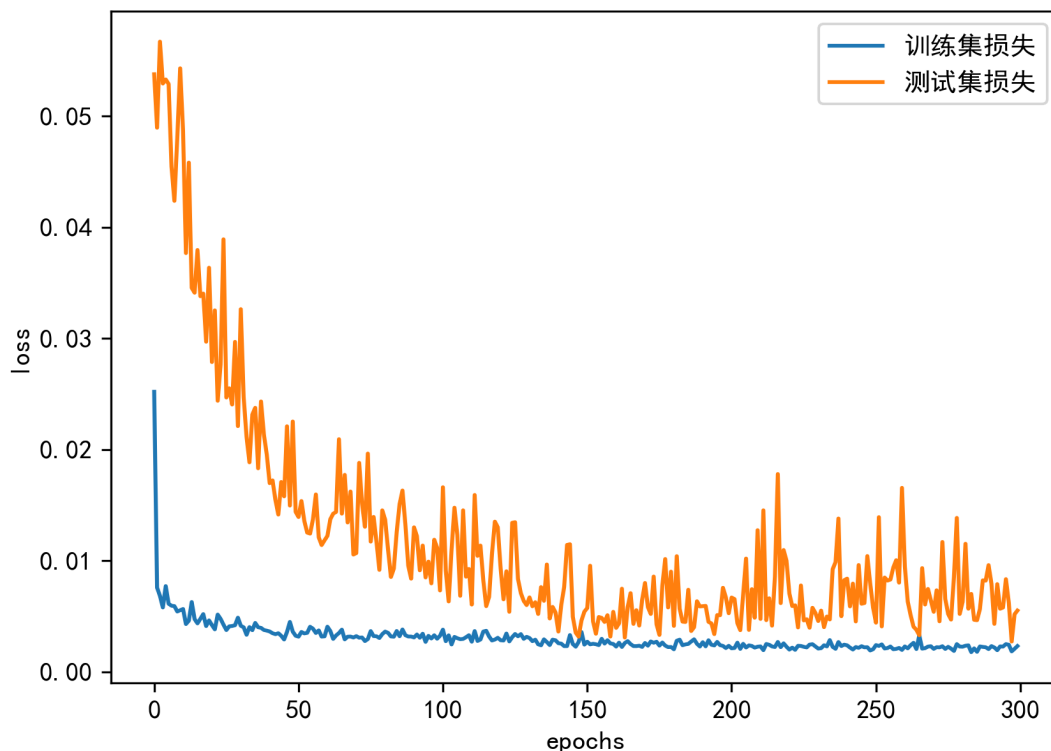


Figure12. Loss graph for 300 iterations

图 12. 迭代 300 次的损失图

`loss` 是训练数据上的损失, 衡量模型在训练集上的拟合能力。`val loss` 是在测试集上的损失, 衡量的是在未见过数据上的拟合能力, 也可以说是泛化能力。模型的真实效果应该用 `val loss` 来衡量。从图中可

得，模型在迭代 125 次时就已经最优了，所以更新到迭代 125 次重新训练，见图 13。

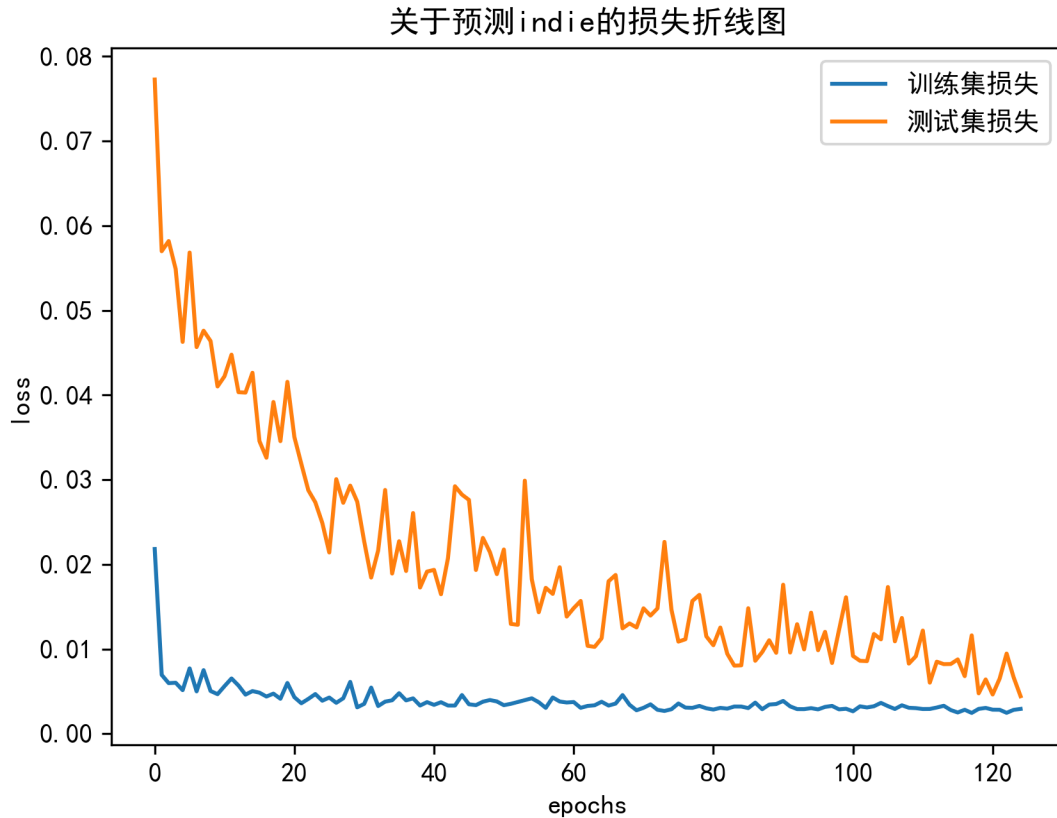


Figure13. Loss graph for 125 iterations  
图 13. 迭代 125 次的损失图

总结，该模型最好的参数选择：batch\_size 为 6；迭代次数为 125。

### 2.3.4. 预测

最后得到 batch\_size 为 6；迭代次数为 125 的模型重新训练，对模型进行预测。如图 14。

观察其结果，可得该效果很好，充分的反映了预测的趋势。则我们可以利用该模型对未来走势的预测。

## 3. 总结

综上，将所有模型计算出的误差(MSE, MAE)进行汇总，这里并没有考虑 SEIRD 模型，因为 SEIRD 模型得到的结果并不能很好反映对未来的预测(太依赖于对参数的取值)，仅用于理论研究更好，而机器学习和 LSTM 得到的结果，见表 2。

Table 2. Error  
表 2. 误差

模型	MSE	MAE
决策树	0.4756	0.2678
线性回归	0.3159	0.1645
支持向量机	0.2829	0.1409
LSTM	0.0122	0.0849

综上, 根据得到的误差分析, LSTM 模型比较好, 所以我们将采用 LSTM 模型对印度的疫情预测, 结果预测出在未来印度的疫情, 会受到新冠疫苗的接种, 应急药品和应急设备设施储备的增加, 疫情的监控和预警这些措施的影响后呈下降趋势, 新增人数会愈来愈少[19] [20], 见图 14。

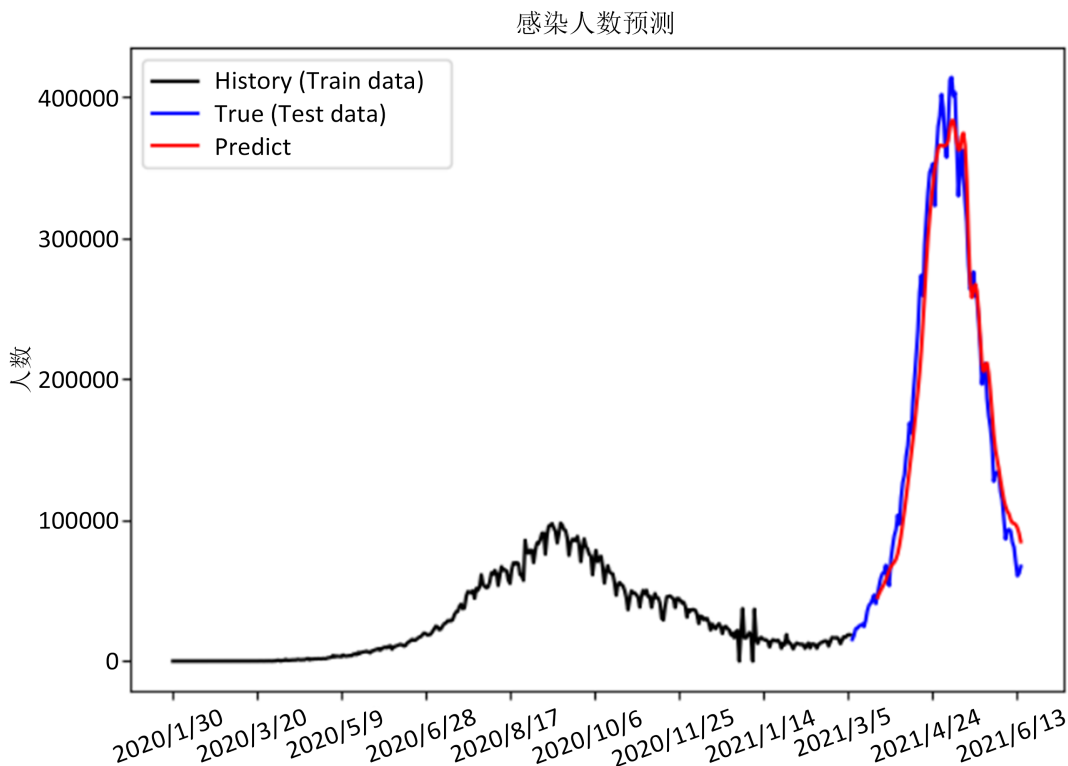


Figure14. LSTM prediction results  
图 14. LSTM 预测结果

## 基金项目

本文收到重庆市大学生创新创业项目(s202211551027)支持。

## 参考文献

- [1] World Health Organization (2021) Coronavirus Disease (COVID-19) Outbreak. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [2] Nishiura, H., Linton, N.M. and Akhmetzhanov, A.R. (2020) Serial Interval of Novel Coronavirus (COVID-19) Infections. *International Journal of Infectious Diseases*, **93**, 284-286. <https://doi.org/10.1016/j.ijid.2020.02.060>
- [3] Kucharski, A.J., Russell, T.W., Diamond, C., *et al.* (2020) Early Dynamics of Transmission and Control of COVID-19: A Mathematical Modelling Study. *The Lancet Infectious Diseases*, **20**, 553-558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
- [4] Huang, L., Zhang, X., Zhang, X., *et al.* (2020) Rapid Asymptomatic Transmission of COVID-19 During the Incubation Period Demonstrating Strong Infectivity in a Cluster of Youngsters Aged 16-23 Years Outside Wuhan and Characteristics of Young Patients with COVID-19: A Prospective Contact-Tracing Study. *Journal of Infection*, **80**, e1-e13. <https://doi.org/10.1016/j.jinf.2020.03.006>
- [5] Li, Q., Guan, X., Wu, P., *et al.* (2020) Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *New England Journal of Medicine*, **382**, 1199-1207. <https://doi.org/10.1056/NEJMoa2001316>
- [6] Grassly, N.C., Pons-Salort, M., Parker, E.P.K., White, P.J. and Ferguson, N.M. (2020) Comparison of Molecular Testing Strategies for COVID-19 Control: A Mathematical Modelling Study. *The Lancet Infectious Diseases*, **20**, 1381-1389. [https://doi.org/10.1016/S1473-3099\(20\)30630-7](https://doi.org/10.1016/S1473-3099(20)30630-7)

- 
- [7] Zhou, F., Yu, T., Du, R., *et al.* (2020) Clinical Course and Risk Factors for Mortality of Adult Inpatients with COVID-19 in Wuhan, China: A Retrospective Cohort Study. *The Lancet*, **395**, 1054-1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)
- [8] Wang, C.J., Ng, C.Y. and Brook, R.H. (2020) Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. *JAMA*, **323**, 1341-1342. <https://doi.org/10.1001/jama.2020.3151>
- [9] Lai, S., Ruktanonchai, N.W., Zhou, L., *et al.* (2020) Effect of Non-Pharmaceutical Interventions to Contain COVID-19 in China. *Nature*, **585**, 410-413. <https://doi.org/10.1038/s41586-020-2293-x>
- [10] Hellewell, J., Abbott, S., Gimma, A., *et al.* (2020) Feasibility of Controlling COVID-19 Outbreaks by Isolation of Cases and Contacts. *The Lancet Global Health*, **8**, 488-496. [https://doi.org/10.1016/S2214-109X\(20\)30074-7](https://doi.org/10.1016/S2214-109X(20)30074-7)
- [11] Kucharski, A.J., Klepac, P., Conlan, A.J.K., *et al.* (2020) Effectiveness of Isolation, Testing, Contact Tracing, and Physical Distancing on Reducing Transmission of SARS-CoV-2 in Different Settings: A Mathematical Modelling Study. *The Lancet Infectious Diseases*, **20**, 1151-1160. [https://doi.org/10.1016/S1473-3099\(20\)30457-6](https://doi.org/10.1016/S1473-3099(20)30457-6)
- [12] Ng, Y., Li, Z., Chua, Y.X., *et al.* (2020) Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore—January 2-February 29, 2020. *Morbidity and Mortality Weekly Report*, **69**, 307-311. <https://doi.org/10.15585/mmwr.mm6911e1>
- [13] Wang, D., Hu, B., Hu, C., *et al.* (2020) Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA*, **323**, 1061-1069. <https://doi.org/10.1001/jama.2020.1585>
- [14] Lauer, S.A., Grantz, K.H., Bi, Q., *et al.* (2020) The Incubation Period of Coronavirus Disease 2019 (COVID-19) from Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, **172**, 577-582. <https://doi.org/10.7326/M20-0504>
- [15] Guan, W.-J., Ni, Z.-Y., Hu, Y., *et al.* (2020) Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, **382**, 1708-1720. <https://doi.org/10.1056/NEJMoa2002032>
- [16] Chan, J.F.-W., Yuan, S., Kok, K.-H., *et al.* (2020) A Familial Cluster of Pneumonia Associated with the 2019 Novel Coronavirus Indicating Person-to-Person Transmission: A Study of a Family Cluster. *The Lancet*, **395**, 514-523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)
- [17] Wang, L., Wong, J.E., Ho, K.F., *et al.* (2020) Global Impact of COVID-19 Pandemic on Carbon Emissions. *Earth's Future*, **8**, e2020EF001659.
- [18] Eikenberry, S.E., Mancuso, M., Iboi, E., *et al.* (2020) To Mask or Not to Mask: Modeling the Potential for Face Mask Use by the General Public to Curtail the COVID-19 Pandemic. *Infectious Disease Modelling*, **5**, 293-308. <https://doi.org/10.1016/j.idm.2020.04.001>
- [19] Pan, A., Liu, L., Wang, C., *et al.* (2020) Association of Public Health Interventions with the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*, **323**, 1915-1923. <https://doi.org/10.1001/jama.2020.6130>
- [20] Zhao, S., Lin, Q., Ran, J., *et al.* (2020) Preliminary Estimation of the Basic Reproduction Number of Novel Coronavirus (2019-nCoV) in China, from 2019 to 2020: A Data-Driven Analysis in the Early Phase of the Outbreak. *International Journal of Infectious Diseases*, **92**, 214-217. <https://doi.org/10.1016/j.ijid.2020.01.050>