COVID-19 Epidemic Prediction Based on Deep Learning

Rui Li[1], Zhihan Zhang[2] and Peng Liu[1,†]

Abstract In this paper, a multi-layer gated recurrent unit neural network (multi-head GRU) model is proposed to predict the confirmed cases of the new crown epidemic (COVID-19). We extract the time series relationship in the data, and the rolling prediction method is adopted to ensure the simple structure of the model and achieve higher precision and interpretability. The prediction results of this model are compared with the LSTM model, the Transformer model and the infectious disease model (SIR). The results show that the proposed model has higher prediction accuracy. The mean absolute error (MAE) of epidemic prediction in most countries (the United States, Brazil, India, the United Kingdom and Russia) is respectively 197.52, 68.02, 200.67, 24.78 and 123.50, which is much smaller than the prediction error of the SIR model, LSTM model and Transformer model. For the spread of the COVID-19 epidemic, traditional infectious disease models and machine learning models cannot achieve more accurate predictions. In this paper, we use a GRU model to predict the real-time spread of COVID-19, which has fewer parameters so that it can reduce the risk of overfitting to train faster. Meanwhile, it can compensate for the transformer model's shortcomings to capture local features.

Keywords COVID-19, deep learning, time series forecasting, gated recurrent unit neural network

MSC(2010) 68T07, 92B20.

1. Introduction

In December 2019, many patients with "pneumonia of unknown cause" appeared in Wuhan, Hubei Province, China. Through the monitoring of related diseases, a series of infection cases of new atypical viral pneumonia were successively discovered. Since the population in China was at the peak of returning to their hometown during the Spring Festival, the new type of viral pneumonia quickly spread to other regions of China and even neighboring countries in a very short period. On 11 February, 2020, the increasingly severe pneumonia epidemic caused concerns in many countries. The World Health Organization named the pneumonia epidemic caused by a new type of coronavirus "COVID-19", or the COVID-19 epidemic for

---

[†]The corresponding author.

Email address: 2030351@tongji.edu.cn (R. Li), zhangsteven2005@163.com (Z. Zhang), geoffreyliu@tongji.edu.cn (P. Liu)

[1]School of Economics and Management, Tongji University, Shanghai 200092, China

[2]Minhang Crosspoint Academy at Shanghai Wenqi Middle School, Shanghai 200240, China

short.

At the early stage of the outbreak of COVID-19 pneumonia, the Chinese government took many intervention measures, including isolating suspected and confirmed patients, closing many transportation channels in Wuhan and timely lengthening the Spring Festival holiday time to avoid population shift, the epidemic had been effectively controlled.

Since March 2020, the epidemic spread rapidly in the United States, the United Kingdom, India and other countries. Up to 5 March, 2022, a total of 443,450,514 COVID-19 cases had been confirmed worldwide with a total of 6,008,173 deaths. Globally, there were 1,724,651 new confirmed cases and 8,889 new deaths in a single day. To prevent and control the global pandemic effectively, it is necessary to understand the dynamics of COVID-19 and predict the infection mode.

Basing and Tay [1], as well as Chitnis, Cushing and Hyman [7], established infectious disease models to understand the infection mechanism and propose corresponding control measures. Sharomi et al. [25] studied transmission and established models to analyze the dynamics of infectious disease. Jajarmi et al. [15] successfully controlled the transmission rate of infectious disease in infants using an improved SIRS model. A similar mathematical model proposed by Baleanu et al. [2,13] successfully helped clinicians better understand the characteristics of the human liver and the transmission of dengue outbreaks. Most of the data-driven methods used in the previous studies [19] are linear methods, and they have ignored the time series in the data, and do not capture the transmission dynamics of novel coronavirus. Statistical models, such as autoregressive moving average (ARIMA), moving average (MA) and autoregressive methods (ARs), mainly rely on assumptions, but it is difficult for these models to predict real-time transmission rates. Benvenuto et al. [8,28] established various statistical and mathematical models to simulate the spread of the current COVID-19 outbreak. However, these models do not fit the given data perfectly in many cases, and have low accuracy in predicting the growth of COVID-19 transmission. James and Tripathi [16] used the concept of derivative to calculate the acceleration of confirmed infection and death cases, and then applied the multivariate linear function and the calculus chain rule of the composite function of the confirmed infection to determine the acceleration of the death function. They fit different ARIMA models for the acceleration of each death function, and found that seasonal changes affect the transmission of the virus. Jia [17] et al. proposed two impulsive systems to describe the impact of multilateral imported cases of COVID-19. Based on the published data, they simulated and analyzed the epidemic trends under different control strategies. R_0 is commonly used to measure i to predict how many people will be infected by an infected person where additional weights are placed on people who have never been infected with the current disease or who have not been vaccinated. If a disease has an R_0 value of 10, an infected person will spread the disease to 10 persons around. Zhang et al., [28] used the R_0 method to determine the infection rate of the new virus on the Diamond Prince cruise ship. However, it is difficult to find the origin of the infectious disease by identifying the patient zero and the people whom they interacted with during the incubation period by using this method. Baleanu et al. [3,14,27] proposed complex nonlinear models to address infectious diseases. While these epidemiological models are good at capturing important components of infectious diseases, some assumptions are needed to make on the parameters. In addition, if these hypothetical parameters do not fit the data perfectly, the accuracy of the models is low.