

Forecasting the Number of Infections of Novel Coronavirus with Deep Learning

Vakada Naveen
VR Siddhartha
Engineering College
Kanuru, Vijayawada

Chunduri Aasish
VR Siddhartha
Engineering College
Kanuru, Vijayawada

Manne Kavya
VR Siddhartha
Engineering College
Kanuru, Vijayawada

Meda Vidhyalakshmi
VR Siddhartha
Engineering College
Kanuru, Vijayawada

ABSTRACT

The novel coronavirus has infected thousands of people across the globe and killed hundreds of those infected. It causes an illness called COVID-19. The virus that causes this serious illness is SARS-CoV-2. There number of infections have increased gradually as more and more people are being infected. The epicentre of the outbreak has been under lockdown since the inception of the outbreak. In this paper a deep neural network has been used to forecast the number of coronavirus infections. Various deep learning techniques have been used to forecast the number of infections. The data related to the infections are collected on a daily basis since the beginning of the outbreak and fed to the deep neural network for training. Once the model is trained to achieve the desired accuracy, it is used to forecast the number of infections.

General Terms

Machine learning, deep learning, epidemic modelling ,time series forecasting.

Keywords

Coronavirus, SARS-COV-2, forecasting, deep learning, neural networks, epidemic predictions.

1. INTRODUCTION

In[1] December 2019 a group of people reported pneumonia illness with similar symptoms in the hubei province of wuhan, china. Soon the medical tests revealed that it is caused by a virus similar to SARS-COV-1. These people who reported the symptoms have visited the huanan seafood market, which is believed to be the epicentre of the outbreak. The seafood market was closed immediately by the Chinese authorities.

The virus started infecting hundreds which doubled and trebled into thousands by the end of January 2020. The virus can transmit from one person to another person by droplets of cough or sneeze. The symptoms of the infected people include fever, cough and shortness of breath. Human to human transmission of the virus is confirmed during early January of 2020. The World Health Organization (WHO) declared a global health emergency on January 30. This virus had spread very fast and infected approximately 83,000 people until February 28th 2020. It has also killed 2,850 people until the end of February.

There is no cure or standard treatment for this disease right now. Researchers across the globe are working to find a cure for the disease. The best way to prevent further infections is to stay isolated from others and wash hands regularly. The incubation period of the disease lies between three to fourteen days. The disease has mostly affected elderly people who have previous medical conditions [2].

Currently, the genome sequence of the SARS is under study. The genome sequence information can be used to make

antiviral drugs and vaccines for the deadly disease [3]. The genome sequence also revealed that this novel coronavirus is very similar to the previously known human coronavirus. Hence the researchers are working on new methods to find the antibodies for the virus and vaccine to prevent it.

Slav W. Hermanowicz [4] , in his paper used a simple logistic growth model to predict the number of Coronavirus cases. He used three sets of data to predict the cases. The model predicted maximum cases of 21,000, 28,000 and 35,000 in future. By linearization the time he got 65,000 cases, a very high number of cases. But the data they have used for prediction is until February 6th. After that date the cases still showed a linear trend and increased to 75,000 approximately by the end of February. He also estimated that the number of cases will stabilize in mid-February.

Xinmiao Fu [5] et.al , fitted Boltzmann function to the data. They have also estimated the number of cases in Hubei province of china as 79589 (95% CI 71576, 93855). They also suggested that the number of newly recorded cases will be at minimal by the end of February. The results, they have predicted are close to the actual number on 15th and 16th of February. But their results cannot be guaranteed because of the uncertainty of the data.

Igor Nesteruk [6], used a statistical approach to predict the number of infected people and the number of victims. They have used the SIR model for estimating the numbers. The model is built using linear equations. They have used statistical approaches to find out the parameters of the linear regression model.

Javier Gamero et.al [7], have also used statistical methods to estimate the number of infections. They have tried to fit a quadratic model and a linear model to the data to obtain the results. They also provided a visual representation of the epidemics of the data available. They concluded that the impact of the disease is moderate.

In this paper , the data is fed to a deep neural network and predictions are made using the neural network. The architecture of various models that are used for prediction is also presented. The techniques and modes that are previously used to forecast time series data is discussed in the next few paragraphs.

Brian S Freeman [8] et al., in his paper predicted the air pollution using deep learning techniques. The dataset that has been used to train the model consists of the hourly data of the air quality as well as meteorological data. The deep learning model used is recurrent neural network which is used to forecast the patterns of the air quality and used long short time memory which is used to identify the patterns and dependencies. The model used predicts the ozone levels in the air for upto 72 hours.

Ryo Akitha [9] et al., proposed a deep learning model to predict the financial stock market. They make the stock market predictions based on the past events using paragraph vectors and the long short time memory algorithms. In this paper they use both numerical and the textual information to make predictions and they found the patterns between the different companies to make predictions. The model make predictions of the stock prices for about fifty companies.

Ahmad Alzahrani [10] et al., in his paper predicted solar irradiance using deep recurrent neural networks. The data that has been taken is real data and it is fed to the deep recurrent neural network which make the predictions. The model can forecast the values of the solar irradiance. The results which are obtained by the deep recurrent neural network have more accuracy when compared to other methods.

Cleo Anastassopoulou et al [11] proposed a data-based analysis, where he used the parameters basic reproduction rate, infection, recovery and mortality rate as parameters to forecast the number of people to be affected in the near future. They have used SIRD model to predict the infection rate. The data available from 11th of January to 29th of February. The problem with the above approach is that, for large scale epidemics, the total number of infected people cannot be known precisely.

K.Roosa[12] et al have suggested methods which have been used previously for prediction during the outbreaks of SARS, ebola, influenza, dengue. They generated short term forecasts using three phenomenological models. They have predicted for short periods of time which may have irregularities so cumulative trajectory of epidemic has been used for prediction. 5, 10 and 15 day ahead predictions were generated.

2. METHODOLOGY

The dataset consists of two features. One feature is the number of days since the inception of the epidemic outbreak. The other feature consists of the number of infected persons around the world. This size of the dataset is 32 instances. It contains the infection numbers starting from 22nd january of 2020 to 22nd february 2020.

In this paper, various architectures for the design of deep neural networks to predict the number of infections will be mentioned. Five different deep neural networks are made to forecast the infections. Each neural network has a different number of layers or a different number of neurons in each layer. The hyperparameters like the learning rate of an optimizer are tuned to improve the neural network. The architecture of the neural network, which is named model-1 is shown in figure 1.

Model-1 has a total of 317,153 parameters. This neural network consists of five layers including the input and output layer. There are three hidden layers in this neural network consisting of 750,350 and 150 neurons in each layer. The model is trained for a total of 100 epochs. Nadam optimizer is used in the model. Keras module has been used to build the neural network with less effort. This model has predicted the number of cases for 14th march 2020 as 157,760 which is close to the actual number of 156,653.

Similarly the architectures of the other four models that could be used to forecast the number of infections are shown in figure 2, figure 3, figure 4 and figure 5 respectively. The loss function used for all the models is root mean square error functions. Mean absolute error is used as a metric to measure the accuracy of the model.

All these five neural networks are trained for 100 hundred epochs. The learning rate parameter is tuned between 0 and 1 to improve the accuracy. The forecastings of each possibility is recorded and the best set of parameters for each type of model is found out.

The data available until now is not sufficient to forecast the numbers precisely. The epidemic is at the beginning stage and the number of cases might increase largely depending on various factors like the type of country where it is spreading more, the climate of the region and the response of the regional administrative bodies to the situation. This model predicts that the numbers of infections will rise atleast for another few weeks.

3. RESULTS AND ANALYSIS

The number of infections predicted for each model with various combinations of parameters is shown in table1, table2, table3, table4 and table 5. Model 1 has actually predicted the numbers pretty close to the actual number of cases for 14th march 2020. All these tabulated values shown are predicted for the date 14th march 2020 which is 52 days from the inception of the epidemic disease.

Table 1. Results for model 1

Learning Rate	Forecasted value	Mean Absolute error
0.1	146640	4793
0.2	145382	4894
0.01	157760	4709
0.02	164976	6368
0.05	146664	4787
0.001	148402	4765

Table 2. Results for model 2

Learning Rate	Forecasted value	Mean Absolute error
0.001	150532	5991
0.0001	144312	4897
0.0005	149931	4773
0.00005	146955	4785
0.000001	146665	4787
0.000005	146620	4788

Table 3. Results for model 3

Learning Rate	Forecasted value	Mean Absolute error
0.05	156902	4705
0.001	145496	4857
0.0001	146658	4788
0.0005	148722	4765
0.00001	146695	4788

Table 4. Results for model 4

0.05	146705	4785
0.001	147179	4781
0.0001	146578	4792
0.0005	146401	4791

Table 5. Results for model 5

Learning Rate	Forecasted value	Mean Absolute Error
0.05	144359	4929
0.001	145938	4795
0.005	160707	4771
0.0001	146566	4794

4. ARCHITECTURE OF THE NEURAL NETWORKS

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_5 (Dense)	(None, 1)	2
dense_6 (Dense)	(None, 750)	1500
dense_7 (Dense)	(None, 350)	262850
dense_8 (Dense)	(None, 150)	52650
dense_9 (Dense)	(None, 1)	151

Total params: 317,153
Trainable params: 317,153
Non-trainable params: 0

Figure 1. The architecture of the neural network model-1

Model: "sequential_2"

Layer (type)	Output Shape	Param #
dense_13 (Dense)	(None, 1)	2
dense_14 (Dense)	(None, 900)	1800
dense_15 (Dense)	(None, 820)	738820
dense_16 (Dense)	(None, 650)	533650
dense_17 (Dense)	(None, 490)	318990
dense_18 (Dense)	(None, 325)	159575
dense_19 (Dense)	(None, 245)	79870
dense_20 (Dense)	(None, 1)	246

Total params: 1,832,953
Trainable params: 1,832,953
Non-trainable params: 0

Figure 2. The architecture of the neural network model-2

Model: "sequential_3"

Layer (type)	Output Shape	Param #
dense_21 (Dense)	(None, 1)	2
dense_22 (Dense)	(None, 500)	1000
dense_23 (Dense)	(None, 250)	125250
dense_24 (Dense)	(None, 150)	37650
dense_25 (Dense)	(None, 1)	151

Total params: 164,053
Trainable params: 164,053
Non-trainable params: 0

Figure 3. The architecture of the neural network model-3

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 1)	2
dense_1 (Dense)	(None, 800)	1600
dense_2 (Dense)	(None, 650)	520650
dense_3 (Dense)	(None, 550)	358050
dense_4 (Dense)	(None, 350)	192850
dense_5 (Dense)	(None, 150)	52650
dense_6 (Dense)	(None, 1)	151

Total params: 1,125,953
Trainable params: 1,125,953
Non-trainable params: 0

Figure 4. The architecture of the neural network model-4

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_7 (Dense)	(None, 1)	2
dense_8 (Dense)	(None, 758)	1516
dense_9 (Dense)	(None, 456)	346104
dense_10 (Dense)	(None, 215)	98255
dense_11 (Dense)	(None, 150)	32400
dense_12 (Dense)	(None, 1)	151

Total params: 478,428
Trainable params: 478,428
Non-trainable params: 0

Figure 5. The architecture of the neural network model-5

5. CONCLUSION

There has been a spike in the number of people infected by novel coronavirus. Time series forecasting models can be used to forecast the number of people infected with COVID-19. Deep learning is very useful to predict time series forecasting using a deep neural network. This paper proposed various architectures of such deep neural networks that could be used to forecast the number of infections. The amount of data available is very less to fully analyze the epidemic. Hence as more new data is added the more accurate the models will predict the numbers. This paper has used the data until february 22 and successfully estimated the number of infections on 14th march 2020. Thus, these type of models can be used to alert the governing bodies and take necessary actions to control the outbreak.

6. REFERENCES

- [1] Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao et al. "A novel coronavirus from patients with pneumonia in China, 2019." *New England Journal of Medicine* (2020).
- [2] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y. and Xing, X., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*.
- [3] Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y. and Cloutier, A., 2003. The genome sequence of the SARS-associated coronavirus. *Science*, 300(5624), pp.1399-1404.
- [4] Hermanowicz SW. Forecasting the Wuhan coronavirus (2019-nCoV) epidemics using a simple (simplistic) model. medRxiv. 2020 Jan 1.
- [5] Fu, Xinmiao, Qi Ying, Tiejong Zeng, Tao Long, and Yan Wang. "Simulating and Forecasting the Cumulative Confirmed Cases of SARS-CoV-2 in China by Boltzmann Function-based Regression Analyses." *medRxiv* (2020).
- [6] Nesteruk I. Statistics based predictions of coronavirus 2019-nCoV spreading in mainland China. MedRxiv. 2020 Jan 1.
- [7] Gamero, Javier, Juan A. Tamayo, and Juan A. Martinez-Roman. "Forecast of the evolution of the contagious disease caused by novel coronavirus (2019-nCoV) in China." *arXiv preprint arXiv:2002.04739* (2020).
- [8] Freeman, Brian S., et al. "Forecasting air quality time series using deep learning." *Journal of the Air & Waste Management Association* 68.8 (2018): 866-886.
- [9] Akita, Ryo, et al. "Deep learning for stock prediction using numerical and textual information." *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*. IEEE, 2016.
- [10] Alzahrani, Ahmad, et al. "Solar irradiance forecasting using deep neural networks." *Procedia Computer Science* 114 (2017): 304-313.
- [11] Anastassopoulou, Cleo, Lucia Russo, Athanasios Tsakris, and Constantinos Siettos. "Data-Based Analysis, Modelling and Forecasting of the novel Coronavirus (2019-nCoV) outbreak." *medRxiv* (2020).
- [12] Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., Yan, P. and Chowell, G., 2020. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5, pp.256-263.