Analysis of Neural Network and Statistical Models Used for Forecasting of Covid-19 Cases

T.A. Makarovskikh[1], M.S.A. Abotaleb[1]

[1]*South Ural State University, Lenina, 76, Chelyabinsk, Russia, 454080*

Abstract

More than a year has passed since the Covid-19 epidemic began and no one has been able to forecast the infection cases of the disease with high accuracy. Nowadays a lot of studies are devoted to finding out the pattern of spread of Covid-19 infection and forecast cases of infection. But models used in those studies have big errors in forecasts, and this made it more challenging to discover the pattern of spread of Covid-19. The choice of appropriate models may vary from country to country. One of the biggest challenges that countries face is the increase in the number of Covid-19 cases. In our paper, we analyze errors in the forecast in the list of the top 10 countries affected by Covid-19 using the following types of models: neural network models (machine learning models, LSTM, BiLSTM and GRU, Neural Networks NNAR Model) and statistical models (nonlinear BATS and TBATS models, Holt linear trend, ARIMA), and SIR model and discover the possible field of application of such algorithms.

Keywords

Covid-19, Epidemic, Machine learning, Neural network, LSTM, BiLSTM, GRU, ARIMA, BATS, TBATS, Holt linear trend, SIR model, Nonlinear models

1. Introduction

At the present time, the world is facing the Covid-19 pandemic. According to the reports of the World Health Organization in January 2021 there are more than a hundred million infections in the world, and more than 2 million people died of this virus. Since the emergence of this epidemic in China, researchers from all over the world are still continuing to reveal the pattern of Covid-19 spreading and forecasting the number of infection cases using different models. These models have major errors in forecasting of Covid-19 infection cases that differ from one model to another. Therefore the important role comes from studying and analyzing errors in forecasts of Covid-19 infection cases and the ways to reduce these forecast errors. In our research we studied the machine-learning models and also linear and non-linear models, and applied them to the list of the top 10 countries suffering from high number of Covid-19 cases.

2. Forecasting of Covid-19 Cases by Different Methods

In our report we suggest our algorithm that allows get a forecast using several linear models of short-term forecasting and the epidemiological SIR model, and also neural network models LSTM, BiLSTM and GRU, and NNAR. We developed the R-script for running our algorithm. This algorithm is extensible, and various modules of forecasting using any other methods can be connected to it and analyzed. Thus, using the considered algorithm scheme, it is possible to create a flexible calling function that allows choice from a variety of implemented methods the one that gets the best result in accordance with a given criterion. Currently, we use MAPE as criterion (the examples of calculating MAPE for Germany and Italy is given in tables 1 and 2). In the future, we may give the user an opportunity to choose the criterion for choosing the best forecasting method.

There are different models for better forecasting for different countries. Complex models like BATS and TBATS requiring more computational time may give good results, but these results are comparable with the results obtained by the classical linear models like ARIMA and Holt's model. So,

the use of complicated models, especially those taking into account the floating seasonality, is not advisable in common. As for SIR model that allows to get short-time forecasts with greater error than linear models, it needs lots of computations to calculate the correct and always changing parameters. This is due to the fact that the parameters of the model are selected on the basis of expert assessments and calculations, taking into account many factors that affect the spread of the epidemic. At the moment, it is still difficult to conclude whether it is advisable to discard meaningful information about the spread of the epidemic and use a numerical vector to search for a pattern. But in accordance with the available data, it can be argued that the use of such anonymized vectors makes it possible to obtain short-term forecasts with an error of less than 1%.

Table 1
MAPE for neural network model on testing data (From 2021-01-04 To 2021-01-10)

Country	NNAR(1,5)	LSTM Model	BI LSTM	GRU Model	Average
Spain	0.303 %	0.70 %	0.26 %	1.19 %	0.61325 %
Germany	1.933 %	1.06 %	0.66 %	6.07 %	2.43075 %

Table 2
MAPE for linear models and SIR model on testing data (From 2021-01-04 To 2021-01-10)

Country	BATS	TBATS	Holt's linear Model	ARIMA	SIR	Average
Spain	0.725 %	0.31 %	0.827 %	0.398 %	3.772 %	1.2064 %
Germany	0.656 %	0.711 %	0.816 %	0.623 %	6.808 %	1.9228 %

3. Conclusion

Hence, using the developed R script we can analyze errors in the forecast for any time series using neural network, and statistical models, and also epidemiological SIR model. Particularly, this allows to make a rapid short-time forecast for the current data on Covid-19 spreading for any country. Analyzing the average errors of forecasts for Covid-19 cases by machine learning models we obtain that in common they are greater than the average errors of forecast by statistical models. Hence, machine learning models are not effective for forecasting these time series. The study of other models and the ways of improving the quality of the forecasts is the topic of future researches.

4. Acknowledgements

The work was supported by the Ministry of Science and Higher Education of the Russian Federation (government order FENU-2020-0022).

5. References

[1] Abotaleb, M.S.A. Predicting COVID-19 Cases using Some Statistical Models: An Application to the Cases Reported in China Italy and USA // Academic Journal of Applied Mathematical Sciences. – 2020. – Vol. 6(4). – P. 32-40.
[2] Hyndman, R.J. Forecasting: principles and practice / R.J. Hyndman, G. Athanasopoulos // OTexts, 2018.
[3] Chimmula, V.K.R. Time series forecasting of COVID-19 transmission in Canada using LSTM networks / V.K.R. Chimmula, L. Zhang // Chaos, Solitons & Fractals. – 2020. – Vol. 135. – P. 109864.
[4] Kırbaş, İ. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches / İ. Kırbaş, A. Sözen, A.D. Tuncer, F.Ş. Kazancıoğlu // Chaos, Solitons & Fractals. – 2020. – Vol. 138. – P. 110015.

[5] Shastri, S. Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study / S. Shastri, K. Singh, S. Kumar, P. Kour, V. Mansotra // Chaos, Solitons & Fractals. – 2020. – Vol. 140. – P. 110227.
[6] Shahid, F. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM / F. Shahid, A. Zameer, M. Muneeb // Chaos, Solitons & Fractals. – 2020. – Vol. 140. – P. 110212.