doi 10.18699/bgrs2024-2.3-03

The forecasting of the COVID-19 spread in Russian Federation regions based on conditional generative adversarial network

Krivorotko O.I.*, Zyatkov N.Yu.
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
** o.i.krivorotko@math.nsc.ru*

Key words: deep learning; GAN; generative AI; machine learning; epidemiology; data processing

Motivation and Aim: COVID-19 caused by a novel coronavirus has continued to pose as a serious public health risk in 2024. The mutation of the virus has reached seasonal frequency, which requires additional resources of the medical system [1]. Despite global efforts to employ several health care strategies for minimizing the impact of the coronavirus on the community, there is still a great need to understand the dynamics of the virus as it transmits from human to human. Mathematical models with computational simulations have been very effective tools that help such global efforts to understand the dynamics of the disease, to estimate key transmission parameters and to make further improvements for controlling this disease [2-5]. The epidemic parameter estimation based on inverse problem approach using measurements about disease propagation such as diagnosed, severe, critical, vaccination and dead cases in fix time [3]. To forecast the spread of the COVID-19 epidemic, it is necessary to consider the socio-economic characteristics that influence the dynamics of the disease (full and/or partial lockdown, migration, emergence of a new strain, testing, vaccination, etc.). These characteristics affect changes in morbidity parameters, which makes the solution to the inverse problem non-unique and/or unstable to measurement errors.
Enough data on the dynamics of COVID-19 in the world since 2020 allows the use of statistical analysis and machine learning methods to build forecasts of epidemic time series. Classical machine learning and time-series approaches do not allow for probabilistic forecasts or have strong assumptions on the form of the future distribution of the target variable [6].
Methods and Algorithms: We apply Condition Generative Adversarial Networks (CGANs) [7] for probabilistic forecasting of the spread of COVID-19 in Saint-Petersburg using epidemic time series, such as new tested, diagnosed, hospitalized, critical (requiring a ventilator), immunity cases, deaths, self-isolation index in considered region [8] as well as new diagnosed COVID-19 cases in the World and its combinations [9] (architecture is presented in Fig. 1). CGAN is an unsupervised deep learning algorithm consists in two neural networks: the generator G generates new samples z of data close to the true data x and the discriminator D distinguishes generated samples of data from the true samples x with condition y. CGAN based on loss function

$$L(D,G) = E_{x\sim p_r(x)}[log\, D(x|y)] + E_{z\sim p_z(z)}\left[log\left(1 - D\big(G(y)\big)\right)\right].$$

We employ the ForGAN [10] architecture, which combines GANs with recurrent neural networks to generate probabilistic forecasts. During training, the generator tries to learn the distribution (under a given condition) of the original data. Then, after training the CGAN, the generator can build an arbitrarily large number of forecast trajectories (generate future data under a given condition from the learned distribution).

We prepare $M = 15$ time series epidemic data (above conditions y) for training and testing by the sliding time window method in the form of fix-target value pairs and divide them into training and test data. Sliding window width $L = 14$, forecast value $K = 5$. We use the previous 14-day data of COVID-19 in Saint-Petersburg to make a forecast for the next 5 days for the newly diagnosed cases, as well as the confidence interval of this forecast. The first step consists in data processing $f(t)$ based on logarithm transformation $f_{new}(t)=ln(f(t)+1)-ln(f(t-X)+1)$, $X=3,7,14$ days, interpolation, outlier removal and smoothing average [6] (see Fig. 2 for example).

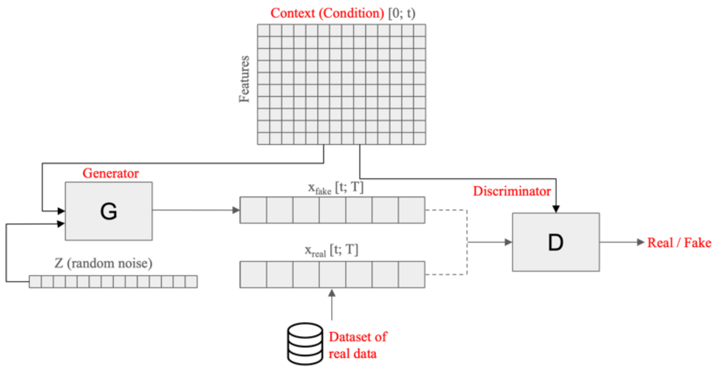

Fig. 1. Architecture of Conditional Generative Adversarial Network

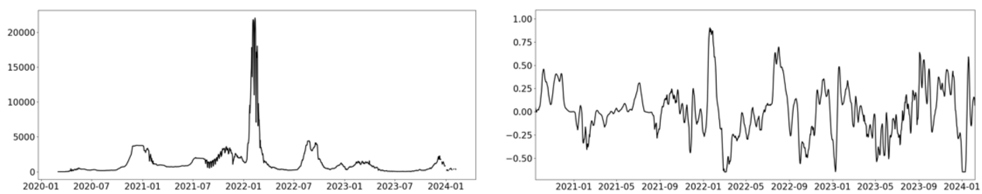

Fig. 2. Number of new diagnosed COVID-19 cases in Saint-Petersburg (left) and its 7-days logarithmic transformation (right)

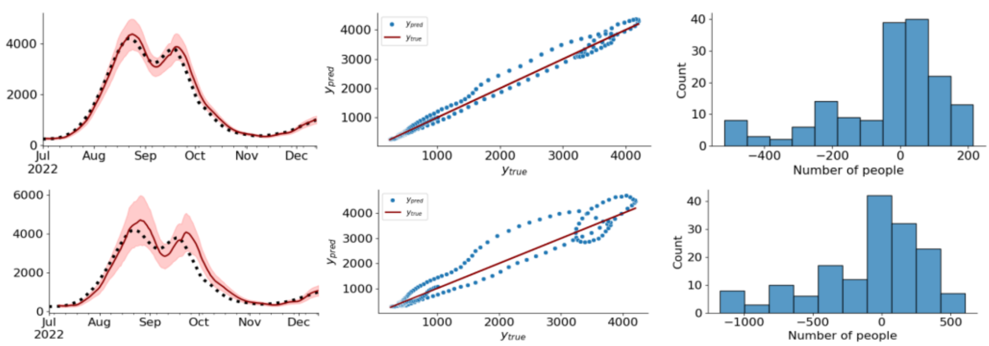

Fig. 3. First column: black dashed line – the real number of new COVID-19 diagnosed cases in St. Petersburg on day t+1 (the first row) and t+5 (the second row); light red area – confidence interval of the forecast on day t+1 and t+5, generated by G; the red line is the average of 10,000 generated forecasts. Second column: the red line is dependence of true forecasts on day t+1 and t+5 on themselves; blue dots – dependence of generator forecasts on true forecasts. Third column: histograms of absolute G forecast errors versus true forecasts

We will feed the discriminator with the target data dataset in order to train it to distinguish real forecasts from fake ones generated by the generator. For hints, we will also introduce a condition (data context) of size *LxM* in both *G* and *D*. *G* generates a prediction based on the white noise and the condition. *D* receives either a false or true prediction and a condition as input. Based on this, the *D* classifies whether the prediction is true or false.

Results: Predictions of newly diagnosed cases of COVID-19 in St. Petersburg were built for 1 and 5 days ahead based on CGAN and real data from 04.2020 to 02.2024 (Fig. 3). The percentage of successful hits in the confidence interval constructed by CGAN for 3-sigma rule for 1 day prediction is 75 % and for 5 days prediction is 67 %. The root mean square error is less than 15 %.

Conclusion: The prediction accuracy is slightly worse than that obtained from the full-connected neural network with and LSTM layers [6] and SEIR-HCD mathematical model based on the system of seven ordinary differential equations [11]. The amount of additional information for the GAN was less than when constructing a fully connected neural network and a SEIR-HCD model. The test data falls within the resulting confidence intervals constructed by the GAN. For more accurate forecasting, it is planned to optimize the network hyperparameters.

Funding: The study is performed to the Government research assignment for Sobolev Institute of Mathematics SB RAS (project FWNF-2024-0002).

References

1. Li Z., Zhang T. Analysis of a COVID-19 Epidemic Model with Seasonality. *Bull Math Biol*. 2022;84(12):146. doi 10.1007/s11538-022-01105-4
2. Krivorotko O.I., Kabanikhin S.I., Zyatkov N.Yu., Prikhodko A.Yu., Prokhoshin N.M., Shishlenin M.A. Mathematical modeling and forecasting of COVID-19 in Moscow and Novosibirsk region. *Num Anal Appl*. 2020;13(4):332-348. doi 10.1134/S1995423920040047
3. Krivorotko O., Kabanikhin S. Artificial intelligence for COVID-19 spread modeling. *J Inverse Ill-Posed Probl.* 2024;32(2). doi 10.1515/jiip-2024-0013
4. Krivorotko O., Sosnovskaia M., Vashchenko I., Kerr C., Lesnic D. Agent-based modeling of COVID-19 outbreaks for New York state and UK: parameter identification algorithm. *Infect Dis Modell*. 2022;7:30-44. doi 10.1016/j.idm.2021.11.004
5. Petrakova V., Krivorotko O. Mean field game for modeling of COVID-19 spread. *J Math Anal Appl.* 2022;514:126271. doi 10.1016/j.jmaa.2022.126271
6. Krivorotko O.I., Zyatkov N.Y., Kabanikhin S.I. Modeling Epidemics: Neural Network Based on Data and SIR-Model. *Comput Math Math Phys*. 2023;63(10):1929-1941. doi 10.1134/s096554252310007x
7. Wang S., Wang G., Fu Q., Song Y., Liu J. IH-TCGAN: Time-Series Conditional Generative Adversarial Network with Improved Hausdorff Distance for Synthesizing Intention Recognition Data. *Entropy*. 2023;25(5):781. doi 10.3390/e25050781
8. COVID-19 data for Saint-Petersburg: https://github.com/alexei-kouprianov/COVID.2019.ru
9. Vuletic M., Prenzel F., Cucuringu M. Fin-GAN: forecasting and classifying financial timeseries via generative adversarial networks. *Quant Finance*. 2024;24(2):175-199. doi 10.1080/14697688.2023.2299466
10. Koochali A., Schichtel P., Dengel A., Ahmed S. Probabilistic forecasting of sensory data with generative adversarial networks–ForGAN. *IEEE Access*. 2019;7:63868-63880. doi 10.1109/ACCESS.2019.2915544
11. Krivorotko O.I., Zyatkov N.Y. Data-driven regularization of inverse problem for SEIR-HCD model of COVID-19 propagation in Novosibirsk region. *Eurasian J Math Comput Appl*. 2022;10(1):51-68. doi 10.32523/2306-6172-2022-10-1-51-68