Journal of Database Management
Volume 33 • Issue 1

Adversarial Training for Predicting the Trend of the COVID-19 Pandemic

Haishuai Wang, Zhejiang Provincial Key Laboratory of Service Robot, College of Computer Science, Zhejiang University, China*

Jiali Ma, Tianjin Normal University, China

Ziping Zhao, Tianjin Normal University, China

Zhenyi Jia, Shanghai Jiaotong University Affiliated Sixth People's Hospital, China

Zhenyan Ji, Beijing Jiaotong University, China

Jun Wu, Beijing Jiaotong University, China

ABSTRACT

It is significant to accurately predict the epidemic trend of COVID-19 due to its detrimental impact on the global health and economy. Although machine learning-based approaches have been applied to predict epidemic trend, standard models have shown low accuracy for long-term prediction due to a high level of uncertainty and lack of essential training data. This paper proposes an improved machine learning framework employing generative adversarial network (GAN) and long short-term memory (LSTM) for adversarial training to forecast the potential threat of COVID-19 in countries where COVID-19 is rapidly spreading. It also investigates the most updated COVID-19 epidemiological data before October 18, 2020 and models the epidemic trend as time series that can be fed into the proposed model for data augmentation and trend prediction of the epidemic. The model is trained to predict daily numbers of cumulative confirmed cases of COVID-19 in Italy, USA, China, Germany, UK, and across the world. The paper further analyzes and suggests which populations are at risk of contracting COVID-19.

KEYWORDS

Coronavirus, COVID-19, COVID-19 Prediction, Epidemic Trend, Gan, LSTM, Machine Learning, Predictive Model

INTRODUCTION

The novel coronavirus, also known as SARS-CoV-2, is a family of viruses that are named after their spiky crown and was first reported on December 31 2019 in Wuhan, China(Liu et al., 2020; F. Zhou et al., 2020). COVID-19 spread explosively throughout China and other countries of the world, especially in the USA, Italy, Spain, and UK. The virus has shown evidence of human-to-human transmission(Awadasseid et al., 2020; Kucharski et al., 2020). World Health Organization (WHO) has declared COVID-19 as a pandemic on March 11, 2020(Takian et al., 2020). Globally, as of 22 October 2020, a total of 41,104,946 confirmed cases of COVID-19, including 1,128,325 deaths, reported to WHO. Comparing to other global pandemic in the history, such as the 1918 influenza

DOI: 10.4018/JDM.299560 *Corresponding Author

pandemic(Morens & Fauci, 2007), Severe Acute Respiratory Syndrome (SARS)(Seto et al., 2003) and Middle East Respiratory Syndrome (MERS)(Hemida et al., 2013), COVID-19 is known as its 1) higher infectious rate; 2) longer incubation period, which various from 2-14 days after exposure to the virus; 3) higher fatality rate, especially for people with weak immune system or underlying diseases(Sharma, 2020). Moreover, some infected people are asymptomatic but still contagious(Day, 2020; Hu et al., 2020).

Although COVID-19 continues to spread globally, the experience of China indicates the outbreak is controllable if the government employs effective mitigation measures(Kraemer et al., 2020; B.-Z. Li et al., 2020). In addition, given the fact that the virus is highly contagious and there is no curative medicine or treatment yet, most countries formulated a variety of policy to curb the spread of COVID-19, including but not limited to Lockdown, Travel Bans, Testing Requirements, Social Distancing, Mask Wearing and Quarantine. Of note, better understanding of the possible future development of the COVID-19 pandemic will help with spreading control and prevention(Shiau et al., 2021). To date, studies are focusing on treatment options(L. Chen et al., 2020; Gao et al., 2020) and clinical symptoms(Gemmar, n.d.; Tian et al., 2020; Zhao et al., 2020), whereas only a few research have concentrated on long-term control and prevention. Given improved understanding of the pandemic trend may provide valuable suggestions on how to effectively control the epidemic spread and relieve the impacts to global economy and society(M. Li et al., n.d.), it is significant to accurate predict the epidemic trend of COVID-19. To this end, many approaches have been developed to analyze and forecast COVID-19. Those forecasting methods can be roughly categorized into two classes: prediction approaches based on CT scans and chest X-ray images, and estimation approaches based on the number of cases of COVID-19. The first category mainly adopts deep learning algorithms to analyze X-ray or CT scans for COVID-19 diagnosis. A CNN-based detection model was proposed to detect COVID-19 (Narin et al., 2020) and the model is trained using 100 chest X-ray images where half are from case group and half are from control group. The model was evaluated using five-fold cross-validation and achieved 98% classification accuracy. Another framework, COVIDX-Net, was proposed(Hemdan et al., 2020) to assist radiologists in diagnosing COVID-19 patients based on X-ray. COVIDX-Net employs several deep learning models and indicates the VGG19 outperforms other models with an F-score of 91%. In addition, Goze et al.(Gozes et al., 2020) automatically identified COVID-19 patients and examined the disease burden quantification on CT scans based on a deep-learning approach and achieved AUC of 99.6%. Wang et al.(S. Wang et al., 2021) employed a modified network inception model to extract regions of interest (ROIs) of sizes from the CT scans, and obtained 82.9% accuracy.

On the other hand, researchers have implemented machine learning based models to forecast the future trends of COVID-19, i.e., the estimated number of confirmed, recovered and death cases. Several studies have applied either standard or modified epidemiological models to predict the future trends, i.e., SIR (Susceptible-Infected-Recovered)(Abuhasel et al., 2020; Alboaneen et al., 2020; Y.-C. Chen et al., 2020; Sujath et al., 2020; Wangping et al., 2020) and SEIR (Susceptible-Exposed-Infectious-Removed)(Hou et al., 2020; Pandey, n.d.; Xu et al., n.d.; Yang et al., 2020; X. Zhou et al., n.d.), to predict the epidemic trend of COVID-19. On the other hand, machine learning has been employed as an alternative to the epidemiological models for the epidemic trend prediction, such as Logistic model(Ahmed et al., 2020; Kriston, n.d.; Malavika et al., 2020; Malhotra & Kashyap, n.d.; X. Zhou et al., n.d.), FbProphet(Gaur & Gaur, 2020; Kumar & Susan, 2020; Taylor & Letham, n.d.; P. Wang et al., 2020), Cloud framework(Panda, n.d.; Tuli et al., n.d.), Gaussian mixture model(Bhandari et al., 2020; Singhal et al., 2020), etc. Doganer et al.(Doganer & Zhang, 2021) evaluated the usability of Google Trends data to predicte and model the COVID-19 outbreak. They collected the search words related to coronavirus and showed that Google Trends data can be used to build the forecast model for case numbers in the COVID-19 outbreak. Since the pandemic data are sequences collected over a period of time, a number of machine learning based methods for time series analysis(Maleki et al., 2020; H. Wang et al., 2018; H. Wang, Wu, et al., 2019; H. Wang, Zhang, et al., 2019) could be applied