Applied Soft Computing 122 (2022) 108806

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A comparative study for predictive monitoring of COVID-19 pandemic

Binish Fatimah [a,1], Priya Aggarwal [b,1], Pushpendra Singh [c,*], Anubha Gupta [d]

[a] Department of ECE, CMR Institute of Technology, Bengaluru, India
[b] Vehant Technologies Pvt. Ltd., Noida, India
[c] Department of ECE, National Institute of Technology Hamirpur, HP, India
[d] SBILab, Department of ECE, IIIT-Delhi, Delhi, India

ARTICLE INFO

Article history:
Received 10 July 2021
Received in revised form 2 January 2022
Accepted 31 March 2022
Available online 7 April 2022

Keywords:
COVID-19 modeling
Gaussian mixture model
Composite logistic growth function
SIRD model
ARIMA model
Dictionary learning model

ABSTRACT

COVID-19 pandemic caused by novel coronavirus (SARS-CoV-2) crippled the world economy and engendered irreparable damages to the lives and health of millions. To control the spread of the disease, it is important to make appropriate policy decisions at the right time. This can be facilitated by a robust mathematical model that can forecast the prevalence and incidence of COVID-19 with greater accuracy. This study presents an optimized ARIMA model to forecast COVID-19 cases. The proposed method first obtains a trend of the COVID-19 data using a low-pass Gaussian filter and then predicts/forecasts data using the ARIMA model. We benchmarked the optimized ARIMA model for 7-days and 14-days forecasting against five forecasting strategies used recently on the COVID-19 data. These include the auto-regressive integrated moving average (ARIMA) model, susceptible–infected–removed (SIR) model, composite Gaussian growth model, composite Logistic growth model, and dictionary learning-based model. We have considered the daily infected cases, cumulative death cases, and cumulative recovered cases of the COVID-19 data of the ten most affected countries in the world, including India, USA, UK, Russia, Brazil, Germany, France, Italy, Turkey, and Colombia. The proposed algorithm outperforms the existing models on the data of most of the countries considered in this study.

1. Introduction

In early December 2019, cases of the coronavirus disease (COVID-19) originated in Wuhan city by a Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [1,2]. Within a short span of time, this virus quickly spread to a large population all over the world. It was declared an epidemic by the World Health Organization (WHO) on 11th March 2020. This disease is highly contagious and has infected millions of people globally. The number of deaths reported globally as of 15th May 2021 is more than 3.5 million. It has hugely affected economic activities [3] and plunged millions into poverty. In countries such as USA, Brazil, Italy, and India, the rapid increase in the number of cases caused tremendous stress on the health care system. Its spread has been contained to some extent in various countries by using partial and complete lock-downs, maintaining social distancing, and imposing quarantine for the infected people. For the timely implementation of these measures, a mathematical understanding of the future trend of the spread of disease is required. This can help the authorities announce control measures at an appropriate time. Thus, an accurate forecast of COVID-19 cases is extremely important to control its rapid spread and hence, ensure the safety of the general public.

Researchers across the world have proposed various data-driven methods to forecast COVID-19 data, which has been a difficult and challenging task [4,5]. Predicting or forecasting refers to estimating future cases on the basis of present and past data. It is carried out majorly using two popular approaches. The first approach includes compartmental models such as SIR, SIRD, SEIR models [6] and the second is based on time-series learning methods such as curve-fitting [7,8], autoregression [9,10], and deep learning on time-series data [11,12].

Compartmental models are the traditional methods of forecasting infectious diseases [13]. In these models, the spread of infectious diseases is simulated by stochastic differential equations that describe interactions between different compartments of the population (e.g. susceptible, infectious, and recovered). This approach majorly includes Susceptible–Infected–Removed (SIR) [14], Susceptible–Infected–Removed–Death (SIRD) [15,16] and Susceptible–Exposed–Infected–Removed (SEIR) models [17]. Hybrid models designed using compartmental models and deep learning frameworks have also been proposed recently [18]. Compartmental models are based on the assumption that the chance of an infected person to infect another susceptible person is

---

[*] Corresponding author.

E-mail addresses: binish.fatimah@gmail.com (B. Fatimah), priyaa@iiitd.ac.in (P. Aggarwal), pushpendrasingh@iitkalumni.org (P. Singh), anubha@iiitd.ac.in (A. Gupta).

[1] Equal contribution.

https://doi.org/10.1016/j.asoc.2022.108806
1568-4946/