DSTA 2021 dys-ta.com
16th INTERNATIONAL CONFERENCE
Dynamical Systems – Theory and Applications
December 6-9, 2021, On-line

Predicting Future from COVID-19 Time Series Data Using Polynomial Extrapolation

AJAY MANDYAM RANGARAJAN[1*], JEYASHREE KRISHNAN[2], ANDREAS SCHUPPERT[3]

1. AICES Graduate School, RWTH Aachen University, Aachen, Germany [0000-0002-1209-7909]
2, 3. Joint Research Center for Computational Biomedicine (JRC-Combine), RWTH Aachen University, Aachen, Germany [0000-0003-3355-0415], [0000-0003-3783-6605]
* Presenting Author

Abstract:

Coronavirus pandemic is an ongoing global crisis that requires intervention measures to help curb the spread and provide timely warnings to hospitals on expected cases. Research in prediction on this area is largely focussed on modeling the dynamics (such as SIS model), or rely on data-driven methods such as deep learning or graph based machine learning techniques. However, much of the data coming out of the pandemic, particularly at region level, may be insufficient to use these approaches to predict the immediate future. Here we propose a reliable numerical method based on polynomial projection and extrapolation that can estimate the true number of cases everyday based on the past data for very small times series datasets. The proposed method relies on smoothening out small-scale variations using moving averages, followed by cubic spline interpolation to move from discrete to continuous representation of the data followed by polynomial projection and extrapolation for prediction. The key to this numerical extrapolation relies on chunking the time series into train, benchmark and predict slices wherein different polynomial orders are allowed to train, the best fit is picked during benchmark (wherein true data is used as reference) followed by moving prediction. This method operates on fewer than 100 data points to train, fit and predict the immediate future with a dynamic moving window method that boosts the robustness of the prediction. This method is in principle, applicable for learning on any other (small) time series data.

Keywords: Polynomial Projection, Vandermonde Matrix, Chebyshev Nodes, Time Series Extrapolation, Small Data

1. Introduction

The ongoing Coronavirus (SARS-Cov-2) pandemic is now a global crisis that has already caused more than 3.3 million deaths worldwide (as of 20th May 2021), with health implications to tens of millions of people and economic fallout around the world. There has been a spike in research related to the virus and in particular, in the areas of modeling epidemic spreading and prevention. Apart from the well-established epidemic models, deep learning based methods and graph based machine learning methods have been commonly used to develop heuristics to stop epidemic or to identify patient zero [4][5][6][7]. However, much of the data coming out of the pandemic including daily new cases or deaths, particularly at region level, are not sufficient to use these approaches to predict the immediate future. It can be expected that future predictions on time series data with few data points is challenging. In this work, we propose a reliable numerical method that can estimate the true number of cases everyday based on the past data. This is helpful for region-wise allocation of appropriate hospital beds. We illustrate the method using the data obtained from the Robert Koch Institute (RKI) for incidences reported in the Aachen area, Germany [3]. The data received from RKI is categorized on the basis of demographics i.e. in terms of age and gender, that facilitates category-wise predictions as well.

2. Results and Discussion

Given cumulative COVID data for cases vs days we can construct cubic splines, Y_0, Y_1, Y_2 (dotted lines in Fig. 1(A)) between raw data points y_0, y_1, y_2, y_3 (blue dots in Fig. 1(A)) respectively. For m points, we get $m-1$ splines. Such a spline reconstruction ensures continuity in the function and first and second derivatives [1]. Let such a polynomial representing the function be given by, $T_k(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_k x^k$. For the polynomial projection we require a set of $k+1$ points, $\bar{x} = [x_0, x_1, \ldots, x_k]$ where the value of the polynomial is *exactly* the same as the function i.e. $T_k(x_i) := f(x_i)$. Substituting the set of points into polynomial equation on can construct $(k+1) \times (k+1)$ system i.e. the *Vandermonde matrix*. The Vandermonde matrix could be ill-conditioned for high k. One could also encounter the issue of Runge's phenomenon if one uses equidistributed datapoint. Thus we use the *Chebyshev nodes* to compute the points $\bar{x}$ [2]. Using these nodes ensures that, $\lim_{k\to\infty} (\max |f(x) - T_k(x)|) < \infty$. Note that Chebyshev nodes are computed between $[-1,1]$, so we use an affine transformation between reference domain to physical domain. The given dataset can now be split into train (Fig. 1(B)), benchmark (Fig. 1(C)) and test (Fig. 1(F)) chunks. At this point, the choice of polynomial order, k is unknown for a given range. This is identified by polynomial of best fit in the benchmark chunk by, $d_{\text{test}} = \dfrac{||T_k(x) - f(x)||_{L^2(\text{benchmark})}}{||f(x)||_{L^2(\text{benchmark})}}$. Now the optimum polynomial order, $k*$ is given by, $k* = \text{argmin}_{k\in[3,k_{\max}]} d_{\text{test}}$ i.e. the which minimizes the L2-distance. As we can see from Fig. 1(C), each of the polynomial order predicts differently and allowed to show the future trajectories. Based on the relative error computed from this benchmark set, the top three best predicting polynomial orders were identified. With this, we do a two-day window moving prediction and estimate the number of cases expected in Aachen in the next week to be 2109 as can be seen in Fig. 1(E) (true number of cases being 1675). With this reliable prediction, we make unknown prediction for the near future (Fig 1. (F)). An implementation of this method is available to be published open source.

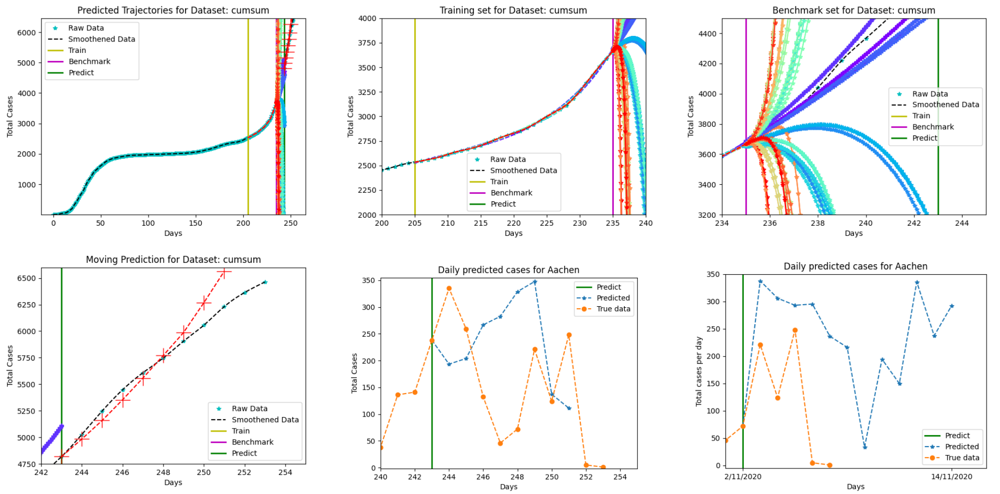

Fig. 1. Predicting Future from COVID-19 RKI Time Series Data for Aachen, Germany using Polynomial Extrapolation, period in days. Curves of different colors indicate predictions from different polynomial order, $k \in [3,20]$: Top Left Subfigure (A) shows the complete time series window including training chunk, benchmark chunk and predict chunk. Top center subfigure (B) shows the training chunk period [205, 235]. Top right subfigure (C) shows the benchmark chunk where the best polynomial order is identified with true data as reference. Bottom left subfigure (D) shows the predict chunk [243, 251] where prediction happens two days at a time. Bottom center subfigure (E) compares the predicted test over a 8 day time interval (2109 cases) to true number of cases over the 8 days (1675 cases). Bottom right subfigure (F) shows the unknown future prediction for the next 12 days

3. Concluding Remarks

We have proposed a method here that promises reliable prediction of future from small time series data such as that of COVID-19 infection statistics obtained locally. The method operates on fewer than 100 data points to train, fit and predict the immediate future. A dynamic moving window method is used at small time steps to make robust predictions. This method is in principle, applicable for learning on any other time series data.

Acknowledgment: The authors would like to thank Pejman Farhadi for the support in handling the RKI data.

References

1. R. H. Bartels, J. C. Beatty, and B. A. Barsky. An Introduction to Splines for Use in Computer Graphics and Geometric Modeling. Morgan Kaufmann Series in Computer Graphics and Geometric Modeling. Elsevier Science, 1995.
2. John P. Boyd and Fei Xu. Divergence (Runge Phenomenon) for Least-Squares Polynomial Approximation on an Equispaced Grid and Mock-Chebyshev Subset Interpolation. In Applied Math. Comput. 210. 1 (2009)
3. Robert Koch Institute (RKI) COVID-19 Pandemic Database, https://www.rki.de/EN/Home/homepage_node.html
4. Gihan Chanaka Jayatilaka et al. Use of Artificial Intelligence on Spatio-Temporal Data to Generate Insights During COVID-19 Pandemic: A Review. MedRXiv (2020)
5. Eli A. Meirom et al. How to Stop Epidemics: Controlling Graph Dynamics with Reinforcement Learning and Graph Neural Networks (2020)
6. Chintan Shah et al. Finding Patient Zero: Learning Contagion Source with Graph Neural Networks. arXiv:2006.11913 (2020)
7. Afshin Shoeibi et al. Automated Detection and Forecasting of COVID-19 using Deep Learning Techniques: A Review. arXiv: 2007.10785v3 (2020)