Adaptive COVID-19 Forecasting via Bayesian Optimization

Nayana Bannur[1], Harsh Maheshwari[2*], Sansiddh Jain[1], Shreyas Shetty[2],
Srujana Merugu[3], Alpan Raval[1]
[1]Wadhwani AI, [2]Flipkart, [3]Independent

ABSTRACT

Accurate forecasts of infections for localized regions are valuable for policy making and medical capacity planning. Existing compartmental and agent-based models [1, 7–11] for epidemiological forecasting employ static parameter choices and cannot be readily contextualized, while adaptive solutions [4, 13] focus primarily on the reproduction number. In the current work, we propose a novel model-agnostic Bayesian optimization approach [3] for learning model parameters from observed data that generalizes to multiple application-specific fidelity criteria. Empirical results demonstrate the efficacy of the proposed approach with SEIR-like compartmental models on COVID-19 case forecasting tasks. A city-level forecasting system based on this approach is being used for COVID-19 response in a few highly impacted Indian cities.

1 INTRODUCTION

The ongoing COVID-19 pandemic and the consequent devastating increase in morbidity and mortality [5] have accentuated the need for robust epidemiological forecasting models. Deployment of such models as part of the public health response requires support for (a) fine-grained contextualization to account for spatio-temporal variations in contact behaviour, lockdown, testing, hospitalization, and reporting policies, (b) multiple models depending on the case count availability (e.g.,age-stratified or testing-based extensions), (c) addressing varying data reliability due to reporting delays, and (d) multiple application use cases with different fidelity requirements (e.g., medical preparedness is tied to accurate $2-4$ week forecasts while long-term policy making might focus on peak estimation). Most existing models [1, 7–11] that use static parameters from domain knowledge and even adaptive likelihood maximization-based methods [4, 13] do not adequately address these requirements.

Problem Statement: For $t \in [0, t_{curr}]$ and region r, given the case count time series $\mathbf{x}(t, r)$ and region metadata $\mathbf{w}(r)$ (e.g. population), forecast $\mathbf{x}_{pred}(t, r)$ for $t \in [t_{curr}, t_{curr} + d]$ s.t. an application specific loss $L(\mathbf{x}_{pred}(:, r), \mathbf{x}(:, r))$ on the forecast period is minimized.

2 PROPOSED SOLUTION

BayesOpt-based Blackbox Learning: For any parametric forecasting model of the form $f_\theta(\mathbf{x}(t), d) = \mathbf{x}_{pred}(t+1 : t+d)$ we optimize $\theta^* = \text{argmin}_\theta \; L(f_\theta(\mathbf{x}(t'), d), \mathbf{x}(t'+1 : t'+d))$ using observations from the period $[t', t'+d]$ for an appropriate loss function $L(\cdot)$. Optimizers such as the hyperopt library [3] can be used.

Uncertainty Estimation: Since certain applications require confidence intervals, the parameter sets (or trials) explored during Bayesian optimization are used to construct a posterior distribution $p(f_\theta(x)|D)$ on the parameter space given data D via a mapping from the observed loss values $L(\cdot)$ and the generative distribution. For instance, in case of exponential families [2], the posterior probability $p(f_\theta(x)|D) \propto \exp(-cL(f_\theta(x), D)$ and c is estimated via validation on a holdout period.

Model Class & Initial Conditions: For practical deployment, we chose SEIR extensions due to their parsimonious nature, flexibility to incorporate testing effects and stratification, and high interpretability. While observed compartments can be readily initialized, the initial values of unobserved compartments (e.g., exposed) are viewed as latent variables and estimated similar to other model parameters, thus also partially accounting for imported cases.

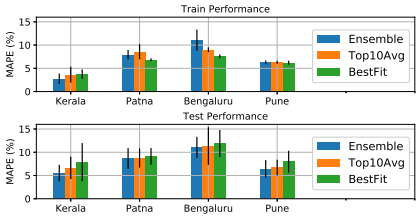

Figure 1: MAPE on case counts from 4 Indian regions.

3 EXPERIMENTS AND RESULTS

We evaluated the efficacy and flexibility of our approach applied to SEIR models on COVID-19 case data [6, 12] of multiple Indian districts for different periods and synthetic data relative to other baselines with relevant choices of loss functions and varying data reliability. Extensive experimentation was performed to identify the parameter search space and the optimal settings for the Bayesian optimization (e.g., the training period) as well as estimate the accuracy for different lead times. For the sake of brevity, we present results with the extended SEIR (Figure 2) model for four regions with train and test periods chosen from July in Figure 1. The forecast variants shown correspond to best fit, average of 10 best trials and an appropriately weighted ensemble of all the trials. The loss function is the average MAPE on all the key case counts. The ensemble-mean provides a stable forecast (test MAPE $< 10\%$).

4 FUTURE DIRECTIONS

Ongoing explorations include alternative methods of estimation of parameters and uncertainty under varying testing and mobility levels, as well as theoretical analysis of control mechanisms for SEIR family models.

* Corresponding author: harsh.maheshwari@flipkart.com.

CoDS COMAD '21, January 02–04, 2021, Bangalore, India

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

5 ACKNOWLEDGEMENTS

This study is made possible by the generous support of the American People through the United States Agency for International Development (USAID). The work described in this article was implemented under the TRACETB Project, managed by WIAI under the terms of Cooperative Agreement Number 72038620CA00006. The contents of this manuscript are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. We thank Anupama Agarwal, Disha Makhija, Mohit Kumar, Sumod Mohan and the COVID modeling team at Wadhwani AI for their contributions to the broader COVID-19 forecasting effort.

REFERENCES

[1] F. Ball, T. Britton, E. Pardoux, C. Larédo, D. Sirl, and V.C. Tran. 2019. *Stochastic Epidemic Models with Inference.* Springer International Publishing. https://books.google.co.in/books?id=LxjBDwAAQBAJ

[2] O.E. Barndorff-Nielsen. 1978. *Information and Exponential Families: In Statistical Theory.* Wiley. https://books.google.co.in/books?id=QxbvAAAAMAAJ

[3] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. 2011. Algorithms for Hyper-Parameter Optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems* (Granada, Spain) *(NIPS'11).* Curran Associates Inc., Red Hook, NY, USA, 2546–2554.

[4] Luís M. A. Bettencourt and Ruy M. Ribeiro. 2008. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. *PLOS ONE* 3, 5 (05 2008), 1–9. https://doi.org/10.1371/journal.pone.0002185

[5] Worldometer Coronavirus Cases. 2020. Coronavirus Cases. https://www.worldometers.info/coronavirus/

[6] Covid19India. 2020. Coronavirus in India: Latest Map and Case Count. https://www.covid19india.org/

[7] Neil Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma M. Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy Okell, Sabine van Elsland, and Azra Ghani. 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. https://doi.org/10.25561/77482

[8] Tiberiu Harko, Francisco Lobo, and M.K. Mak. 2014. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl. Math. Comput.* 236 (03 2014), 184–194. https://doi.org/10.1016/j.amc.2014.03.030

[9] Herbert W. Hethcote. 2000. The Mathematics of Infectious Diseases. *SIAM Rev.* 42, 4 (Dec. 2000), 599–653. https://doi.org/10.1137/S0036144500371907

[10] Lars Lorch, William Trouleau, Stratis Tsirtsis, Aron Szanto, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. 2020. A Spatiotemporal Epidemic Model to Quantify the Effects of Contact Tracing, Testing, and Containment.

[11] Ofir Reich, Guy Shalev, and Tom Kalvari. 2020. Modeling COVID-19 on a network: super-spreaders, testing and containment. *medRxiv* (2020). https://doi.org/10.1101/2020.04.30.20081828

[12] COVID-19 stats in India. 2020. COVID-19 REST API for India. https://api.rootnet.in/

[13] Kevin Systrom, Thomas Vladek, and Mike Krieger. 2020. Rt COVID-19. https://github.com/rtcovidlive/covid-model.

6 APPENDIX

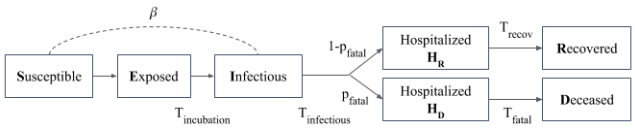

Figure 2: SEIHRD Compartmental Model

The above figure depicts the SEIHRD model with compartments mapping to the key stages of disease progression and parameters corresponding to the infectivity (β), the transition times of various compartments ($T_{incubation}$, $T_{infectious}$, T_{recov}, T_{fatal}), and the probabilities of parallel pathways (p_{fatal}). The observed compartments are initialized from case counts($H_R + H_D$:active, R:recovered, D:deceased) while the unobserved compartments (I:infectious, E:exposed) are handled as latent variables. The equations governing the dynamics are given below.

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta IS}{N} \\
\frac{dE}{dt} &= \frac{\beta IS}{N} - \frac{E}{T_{incubation}} \\
\frac{dI}{dt} &= \frac{E}{T_{incubation}} - \frac{I}{T_{infectious}} \\
\frac{dH_R}{dt} &= \frac{(1-p_{fatal})I}{T_{infectious}} - \frac{H_R}{T_{recov}} \\
\frac{dH_D}{dt} &= \frac{p_{fatal}I}{T_{infectious}} - \frac{H_D}{T_{fatal}} \\
\frac{dR}{dt} &= \frac{H_R}{T_{recov}} \\
\frac{dD}{dt} &= \frac{H_D}{T_{fatal}} \\
N &= S + E + I + H_R + H_D + R + D
\end{aligned}
$$