



Optimising Lockdown Policies for Epidemic Control using Reinforcement Learning

An AI-Driven Control Approach Compatible with Existing Disease and Network Models

Harshad Khadilkar¹ · Tanuja Ganu² · Deva P. Seetharam³

Received: 14 April 2020 / Revised: 4 June 2020 / Accepted: 12 June 2020 / Published online: 24 June 2020
© Indian National Academy of Engineering 2020

Abstract

There has been intense debate about lockdown policies in the context of Covid-19 for limiting damage both to health and to the economy. We present an AI-driven approach for generating optimal lockdown policies that control the spread of the disease while balancing both health and economic costs. Furthermore, the proposed reinforcement learning approach automatically learns those policies, as a function of disease and population parameters. The approach accounts for imperfect lockdowns, can be used to explore a range of policies using tunable parameters, and can be easily extended to fine-grained lockdown strictness. The control approach can be used with any compatible disease and network simulation models.

Keywords Lockdowns · Epidemic Control · Reinforcement Learning

Description of Modelling Approach

In this article¹, we briefly describe an AI-driven approach for generating lockdown policies that are optimised based on disease characteristics and network parameters. This approach is aimed for use by policy-makers who are knowledgeable in epidemiology, but not necessarily well-versed in dynamic systems and control. The approach is designed to be modular and flexible. That is, the underlying reinforcement learning algorithm can work with any compatible disease and network models. Furthermore, the critical characteristics of the models are parameters that can be tuned. The models described in this paper are based

on a commonly used epidemiological model from literature Perez and Dragicevic (2009), as shown in Fig. 1. The parameter values can be tuned for modelling infectious diseases including Covid-19. We use values for Covid-19 computed by Jung et al. (2020).

We also account for network propagation characteristics through tunable parameters such as strictness of lockdowns within network nodes and travel between nodes, including the possibility of leaky quarantine. The probability of disease transmission between people is a macro-level parameter, but it accounts for micro-level effects such as social distancing, mask usage, and weather effects. The network definition is based on node locations and population of each node, with connectivity between each pair of nodes defined by a gravity model Allamanis et al. (2012). The results presented in this paper are based on a randomly generated network with 100 nodes and 10,000 people randomly distributed amongst those

✉ Harshad Khadilkar
harshad.khadilkar@tcs.com; harshadk@iitb.ac.in

Tanuja Ganu
tanuja.ganu@microsoft.com

Deva P. Seetharam
deva.seetharam@gmail.com

¹ TCS Research and IIT Bombay, Mumbai, India

² Microsoft Research, Bangalore, India

³ Independent Systems Researcher, Bangalore, India

¹ This manuscript is an abbreviated version of a broader study Khadilkar et al. (2020).

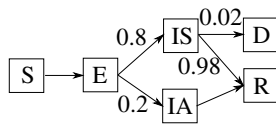


Fig. 1 Disease progression model adapted for Covid-19. S is susceptible, E is exposed (virus in the body but not yet affecting the immune system), IS is infected (showing symptoms), IA is asymptomatic carrier, D is dead, and R is recovered. Note the numbers represent capture probabilities and not rates of change. All parameters can be tuned. We do not consider a transition from Recovered to susceptible states, but this can be added if found to be possible

nodes.² The disease progression is shown in Fig. 2 for a fixed strategy of locking down any node when its symptomatic population exceeds 5% of total, and reopening when it falls below this level. This is a typical method followed in several regions worldwide. While the peak is small, the epidemic lasts for nearly the full year (with high economic cost).

Reinforcement learning approach for computing lockdown policy

Reinforcement learning (RL) works by running a large number of simulations of the spread of the disease while attempting to find the optimal policy for lockdowns Sutton and Barto (2012). The chief requirement is to quantify the cost of each outcome of the simulation. In this study, we impose a cost of 1.0 on each day of lockdown, 1.0 on each person infected, and 2.5 on each death³. A reward is defined as the negative of these costs (higher the reward, lower the cost). The actions asked of the algorithm are binary⁴: at the beginning of every week and for every node, the algorithm must decide whether to keep the node open or lock it down. We use Deep Q Learning Mnih et al. (2015) to train the algorithm. The RL algorithm improves and then saturates in 75 simulations as shown in Fig. 3, for this specific instance. The evolution of infection rates in Fig. 4 (computed through 10 independent runs) shows that the policy has a higher peak than the 5% policy in

Fig. 2, but significantly fewer lockdowns and a shorter epidemic duration. Note also that there are no kinks due to new infections after release.

The key points of novelty in this approach are: (i) we focus neither on epidemiological models nor on prediction of the spread of the disease, but rather on controlling the spread of disease while balancing long-term economic and health costs, (ii) our control approach can work with any disease parameters (not just Covid-19), and with any compatible network data and propagation model (not just for specific geographies), (iii) rather than taking decisions based on simple thresholds such as fraction of people with symptoms, the learnt policies combine several context variables such as rates of new infections to take optimal decisions, (iv) the end-users need to only change input parameters to create policies with their desired characteristics, and (v) the algorithm is not a black box, and sensitivity of the policy to features can be studied. Fig. 5 demonstrates the last claim by considering the sensitivity of decisions to sets of two input features at a time. The first plot shows that the policy recommends lockdowns when the infection rate in the overall population or within a node exceeds 0.2. However, lockdowns can be recommended at much smaller values if both infection rates reach 0.1. A similar trend is shown in the plot on the right, which shows that lockdowns are recommended at much lower infection rates if a node has a large population.

Discussion

The reinforcement learning algorithm is ready for use in conjunction with real-world data sets, epidemiological models and network propagation models. Any of these three aspects can be changed as per user requirements. The algorithm is computationally lightweight, and, running it only requires Python. We have demonstrated its capability of handling nation-scale data Khadilkar et al. (2020). We are open to collaborating with epidemiologists who could benefit from a computational approach to address the spread of communicable diseases.

² The methodology is independent of the population size, and, has been shown to work for real-world networks in an extended version of the manuscript Khadilkar et al. (2020).

³ These numbers can be tuned, and in realistic scenarios, can be made specific to nodes.

⁴ The extended version of the manuscript Khadilkar et al. (2020) demonstrates control with more actions.

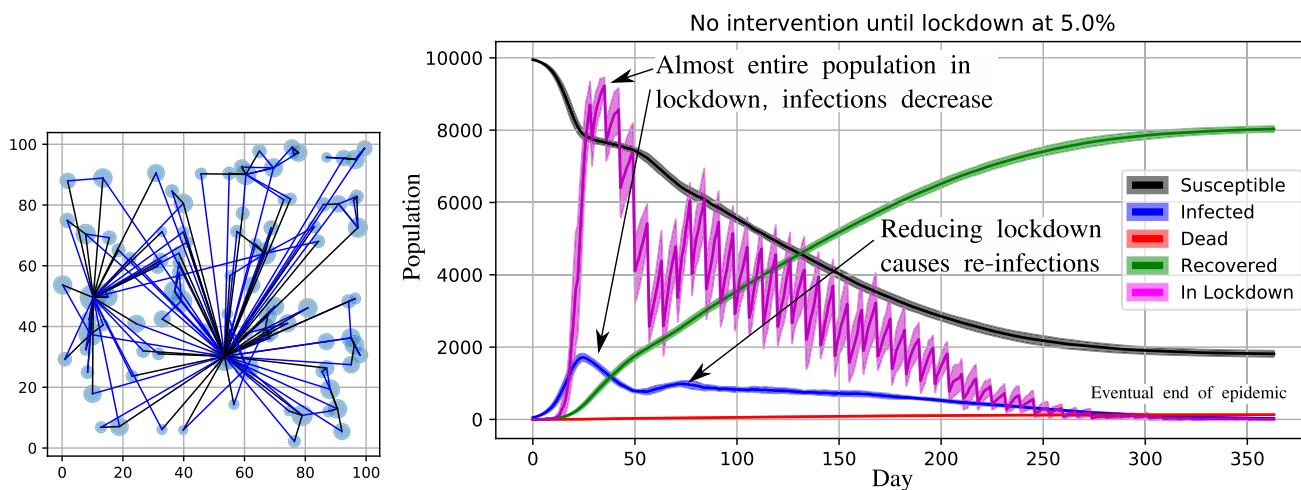


Fig. 2 [Left] Network model, [right] Epidemic progression for the policy of locking down any node when its symptomatic population exceeds 5% of total

Fig. 3 Evolution of reward (objective function) during training, based on (i) duration of lockdowns, (ii) number of people infected, and (iii) number of people dead

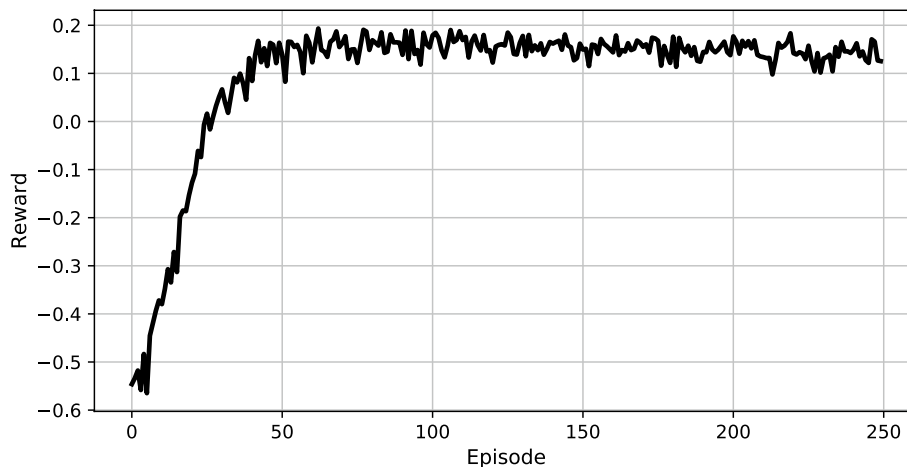
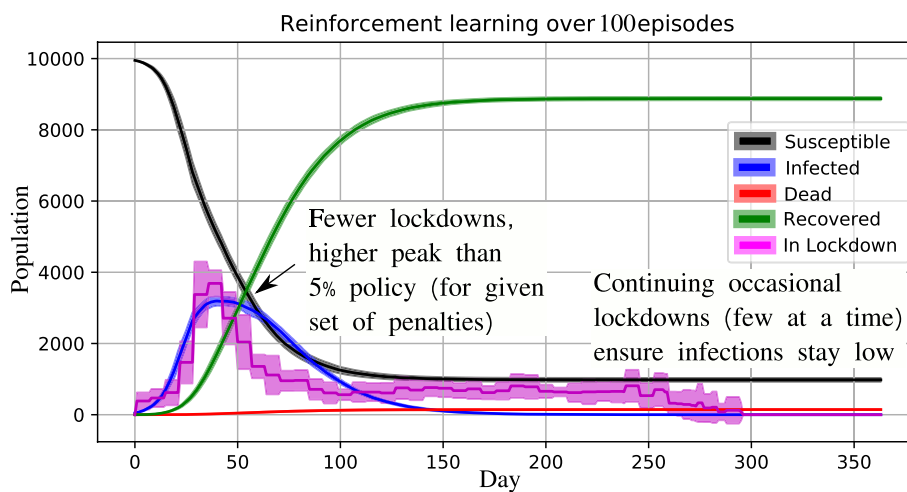


Fig. 4 Epidemic progression using a policy trained by reinforcement learning



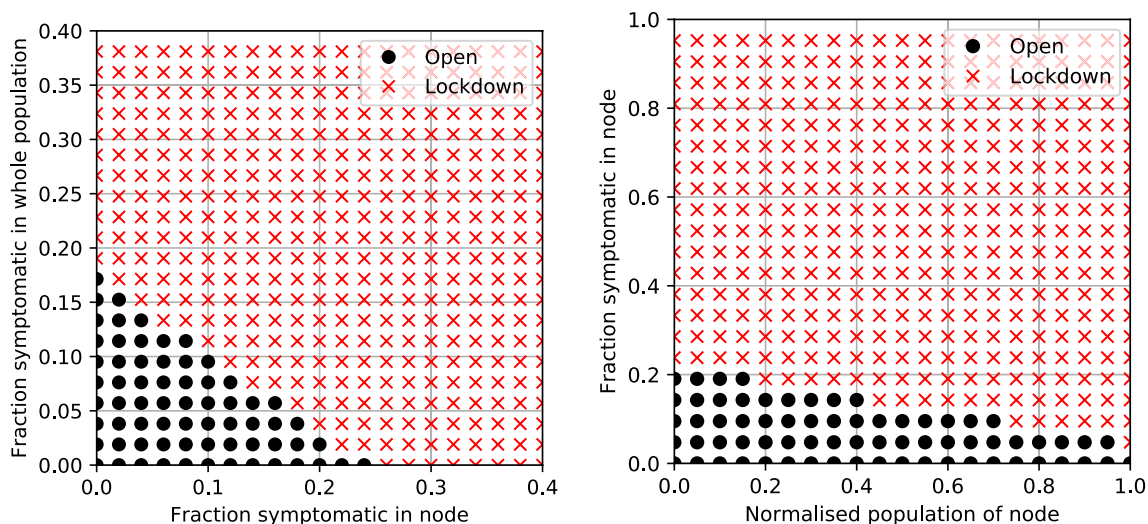


Fig. 5 Policy outcomes as a function of two features at a time

References

- Allamanis M, Scellato S, Mascolo C (2012) Evolution of a location-based online social network: analysis and models, Internet Measurement Conference, pp. 145–158
- Jung S-M, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, Kobayashi T, Kinoshita R, Nishiura H (2020) Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *J Clin Med* 9(2):523. <https://doi.org/10.3390/jcm9020523>
- Khadilkar H, Ganu T, Seetharam DP (2020) Optimising lockdown policies for epidemic control using reinforcement learning, arXiv preprint [arXiv:2003.14093](https://arxiv.org/abs/2003.14093)
- Mnih V, Kavukcuoglu K, Silver D et al. (2015) Human-level control through deep reinforcement learning. *Nature* 518:529–533. <https://doi.org/10.1038/nature14236>
- Perez L, Dragicevic S (2009) An agent-based approach for modeling dynamics of contagious disease spread. *Int J Health Geograph* 8(1):50
- Sutton R, Barto A (2012) Reinforcement learning, 2nd edn. MIT Press, USA

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.