Deep reinforcement learning for large-scale epidemic control

Pieter J.K. Libin[1,2,3], Arno Moonens[1], Timothy Verstraeten[1], Fabian Perez-Sanjines[1], Niel Hens[3], Philippe Lemey[2], and Ann Nowé[1]

[1] Vrije Universiteit Brussel, Brussels, Belgium
{pieter.libin,arno.moonens,timothy.verstraeten,fperezsa,ann.nowe}@vub.be
[2] KU Leuven, Leuven, Belgium
philippe.lemey@kuleuven.be
[3] Hasselt University, Hasselt, Belgium
{pieter.libin,niel.hens}@uhasselt.be

This paper[4] was accepted and presented at the ECML-PKDD 2020 conference.

Epidemics of infectious diseases are an important threat to public health and global economies. The most efficient way to combat epidemics is through prevention. To develop prevention strategies and to implement them as efficiently as possible, a good understanding of the complex dynamics that underlie these epidemics is essential. To properly understand these dynamics, and to study emergency scenarios, epidemiological models are necessary. Such models enable us to make predictions and to study the effect of prevention strategies in simulation. The development of prevention strategies, which need to fulfil distinct criteria (i.a., prevalence, mortality, morbidity, cost), remains a challenging process. For this reason, we investigate a deep reinforcement learning (RL) approach to automatically learn prevention strategies in an epidemiological model. The use of model-free deep reinforcement learning is particularly interesting, as it allows us to set up a learning environment in a complex epidemiological setting (i.e., large state space and non-linear dependencies) while imposing few assumptions on the policies to be learned. In this work, we conduct our experiments in the context of pandemic influenza, where we aim to learn optimal school closure policies to mitigate the epidemic [1].

Pandemic preparedness is important, as influenza pandemics have made many victims in the (recent) past and the ongoing COVID-19 epidemic is yet another reminder of this fact. Contrary to seasonal influenza epidemics, an influenza pandemic is caused by a newly emerging virus strain that can become pandemic by spreading rapidly among naive human hosts (i.e., human hosts with no prior immunity) worldwide. This means that at the start of the pandemic no vaccine will be available and it will take several months before vaccine production can commence. For this reason, learning optimal strategies of non-therapeutic intervention measures, such as school closure policies, is of great importance to mitigate pandemics. To meet this objective, we consider a reinforcement learning approach. However, as the state-of-the-art of reinforcement learning techniques requires many interactions with the environment in order to converge, our first

[4] https://bitbucket.org/ghentdatascience/ecmlpkdd20-papers/raw/master/ADS/sub_616.pdf

contribution entails a realistic epidemiological model that still has a favourable computational performance.

Specifically, we construct a meta-population model that consists of a set of 379 interconnected patches, where each patch corresponds to an administrative region in Great Britain and is internally represented by an age-structured stochastic compartmental model. To conduct our experiments, we establish a Markov decision process with a state space that directly corresponds to our epidemiological model, an action space that allows us to open and close schools on a weekly basis, a transition function that follows the epidemiological model's dynamics, and a reward function that is targeted to the objective of reducing the attack rate (i.e., the proportion of the population that was infected). In this work, we will use "Proximal Policy Optimization" (PPO) to learn the school closure policies.

First, we set up an experiment in an epidemiological model that covers a single administrative district. This setting enables us to specify a ground truth that allows us to empirically assess the performance of the policies learned by PPO. In this analysis, we consider different values for the basic reproductive number R_0 and the population composition (i.e., proportion of adults, children, elderly, adolescents) of the district. Both parameters induce a significant change of the epidemic model's dynamics. Through these experiments, we demonstrate the potential of deep reinforcement learning algorithms to learn policies in the context of complex epidemiological models, opening the prospect to learn in even more complex stochastic models with large action spaces. In this regard, we consider a large scale setting where we examine whether there is an advantage to consider the collaboration between districts when designing school closure policies.

To situate this work in the state-of-the-art, we note that the concept to learn dynamic policies by formulating the decision problem as a Markov decision process (MDP) was introduced in [2]. To our best knowledge, the work presented in this manuscript is the first attempt to use deep reinforcement learning algorithms directly on a complex meta-population model.

To summarize, in this work, we demonstrate the potential of deep reinforcement learning in the context of complex stochastic epidemiological models. As few assumptions are made on the epidemiological model, our new technique has the potential to be used for other epidemiological settings, such as the ongoing COVID-19 pandemic. For future work, it would be interesting to investigate how well these algorithms scale to even larger state and/or action spaces.

References

1. T. C. Germann, H. Gao, M. Gambhir, A. Plummer, M. Biggerstaff, C. Reed, and A. Uzicanin. School dismissal as a pandemic influenza response: When, where and for how long? *Epidemics*, page 100348, 2019.
2. R. Yaesoubi and T. Cohen. Dynamic health policies for controlling the spread of emerging infections: influenza as an example. *PloS one*, 6(9), 2011.