Exploring the Pareto front of multi-objective COVID-19 mitigation policies using reinforcement learning

Mathieu Reymond[1⋆] Conor F. Hayes[2], Lander Willem[3], Roxana Rădulescu[1], Steven Abrams[3], Diederik M. Roijers[1], Enda Howley[2], Patrick Mannion[2], Niel Hens[4], Ann Nowé[1], and Pieter Libin[1]

[1] Vrije Universiteit Brussel, Brussels, Belgium
[2] National University of Ireland Galway, Galway, Ireland
[3] University of Antwerp, Antwerp, Belgium
[4] Hasselt University, Hasselt, Belgium

1 Abstract

Infectious disease outbreaks represent a major challenge [7]. To this end, understanding the complex dynamics that underlie these epidemics is essential. Epidemiological transmission models allow us to capture and understand such dynamics and facilitate the study of prevention strategies through simulation. However, developing efficient mitigation strategies remains a challenging process, given the non-linear and complex nature of epidemics. To address these challenges, reinforcement learning provides a methodology to automatically learn mitigation strategies in combination with complex epidemic models [5]. Previous research focused on optimizing policies with respect to a single objective, such as the pathogen's attack rate, while the mitigation of epidemics is a problem that inherently covers distinct and possibly conflicting criteria (i.a., prevalence, mental health, cost). Therefore, optimizing on a single objective requires that these distinct criteria are somehow aggregated into a single metric. Manually designing such metrics is time-consuming, costly and error-prone, as this non-intuitive process requires repetitive and tedious tuning to achieve the desired behavior [9]. Moreover, taking a single objective approach reduces the explainability of the learned solution, as we cannot compare the learned behavior with alternatives [4].

This challenging process can be circumvented by taking an explicitly multi-objective approach that aims to learn the different trade-offs regarding the considered criteria. By assuming that a decision maker will always prefer solutions for which at least one objective improves, it is possible to learn a set of optimal solutions referred to as the *Pareto front* [4]. This enables decision makers to review each solution on the Pareto front before making a decision, thereby being aware of the trade-offs that each solution implies.

⋆ Corresponding Author. Email: mathieu.reymond@vub.be

In this work[5], we investigate the use of *multi-objective reinforcement learning* (MORL) to learn a set of solutions that approximate the Pareto front of multi-objective epidemic mitigation strategies. We consider the first wave of the Belgian COVID-19 epidemic, which was mitigated by a strict lockdown [12]. When the incidence of confirmed cases was steadily decreasing, epidemiological experts were tasked to investigate deconfinement strategies, to reduce the severe social contact and mobility restrictions.

We consider an epidemiological model that was constructed to describe the Belgian COVID-19 epidemic and was fitted to hospitalization incidence data and serial sero-prevalence data [1]. This model concerns a discrete-time stochastic model that considers an age-structured population. Based on this model, we contribute a novel multi-objective epidemiological reinforcement learning environment (Multi-Objective Belgian COVID environment, MOBelCov), in the form of a multi-objective Markov decision process (MOMDP) [9]. To model different types of non-pharmaceutical interventions, we consider a contact reduction function that imposes a proportional reduction of work (including transport), school and leisure contacts. At each timestep (here one week), our RL agent modulates the contact reduction in these social environments to control the spread of the epidemic.

In multi-objective optimization, the set of optimal policies can grow exponentially with the number of objectives. Thus, recovering them all is a computationally expensive process and requires an exhaustive exploration of the complete state space. To address this problem, we extend *Pareto Conditioned Networks* (PCN), a method that uses a single neural network to encompass all non-dominated policies [8]. As PCN makes no assumptions about the shape of the coverage set, it is particularly well suited for the complex decision problem that we consider, for which the shape of the coverage set is not known a priori.

We learn a coverage set that almost completely dominates the coverage set of the fixed policies generally used in practice. This is most evident in the compromising policies, where one has to carefully choose when to remove social restrictions while at the same time minimizing the impact on daily new hospitalizations. In these scenarios, our algorithm learns policies that drastically reduce the total number of new hospitalizations for the same social burden.

Finally, the methodology that we propose shows promise to address a wide variety of public health challenges, such as balancing the number of lost schooldays with respect to the attack rate of infections in schools [11], the efficacy versus burden of face masks for children [3], contact tracing effort compared to the impact of such policies [12], the impact of antivirals on the epidemic while balancing the likelihood for resistance mutations to emerge [10], to balance the efforts and insights of COVID-19 genomic surveillance [2], and to balance the cost of universal testing and its impact on an emerging epidemic [6].

[5] This work was published in the journal *Expert systems with Applications* and can be freely accessed at `https://doi.org/10.1016/j.eswa.2024.123686`

Bibliography

[1] Abrams, S., Wambua, J., Santermans, E., Willem, L., Kuylen, E., Coletti, P., Libin, P., Faes, C., Petrof, O., Herzog, S.A., et al.: Modelling the early phase of the belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. Epidemics **35**, 100449 (2021)

[2] Chen, Z., Azman, A.S., Chen, X., Zou, J., Tian, Y., Sun, R., Xu, X., Wu, Y., Lu, W., Ge, S., et al.: Global landscape of sars-cov-2 genomic surveillance and data sharing. Nature genetics **54**(4), 499–507 (2022)

[3] Esposito, S., Principi, N.: To mask or not to mask children to overcome COVID-19. European journal of pediatrics **179**(8), 1267–1270 (2020)

[4] Hayes, C.F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Reymond, M., Verstraeten, T., Zintgraf, L.M., Dazeley, R., Heintz, F., Howley, E., Irissappane, A.A., Mannion, P., Nowé, A., Ramos, G., Restelli, M., Vamplew, P., Roijers, D.M.: A practical guide to multi-objective RL and planning (2021)

[5] Libin, P.J.K., Moonens, A., Verstraeten, T., Perez-Sanjines, F., Hens, N., Lemey, P., Nowé, A.: Deep reinforcement learning for large-scale epidemic control. In: Dong, Y., Ifrim, G., Mladenić, D., Saunders, C., Van Hoecke, S. (eds.) ECML. pp. 155–170. Springer International Publishing, Cham (2021)

[6] Libin, P.J., Willem, L., Verstraeten, T., Torneri, A., Vanderlocht, J., Hens, N.: Assessing the feasibility and effectiveness of household-pooled universal testing to control COVID-19 epidemics. PLoS computational biology **17**(3), e1008688 (2021)

[7] Miranda, M.N., Pingarilho, M., Pimentel, V., Torneri, A., Seabra, S.G., Libin, P.J., Abecasis, A.B.: A tale of three recent pandemics: Influenza, hiv and sars-cov-2. Frontiers in Microbiology **13** (2022)

[8] Reymond, M., Eugenio, B., Nowé, A.: Pareto conditioned networks. In: Proceedings of the 21st International Conference on AAMAS (2022) (2022)

[9] Roijers, D.M., Vamplew, P., Whiteson, S., Dazeley, R.: A survey of multi-objective sequential decision-making. JAIR **48**, 67–113 (2013)

[10] Torneri, A., Libin, P., Vanderlocht, J., Vandamme, A.M., Neyts, J., Hens, N.: A prospect on the use of antiviral drugs to control local outbreaks of COVID-19. BMC medicine **18**(1), 1–9 (2020)

[11] Torneri, A., Willem, L., Colizza, V., Kremer, C., Meuris, C., Darcis, G., Hens, N., Libin, P.J.: Controlling SARS-CoV-2 in schools using repetitive testing strategies (preprint) (2021)

[12] Willem, L., Abrams, S., Libin, P.J., Coletti, P., Kuylen, E., Petrof, O., Møgelmose, S., Wambua, J., Herzog, S.A., Faes, C., et al.: The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19. Nature communications **12**(1), 1–9 (2021)