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

Mathieu Reymond[a,*], Conor F. Hayes[b], Lander Willem[c], Roxana Rădulescu[a], Steven Abrams[c], Diederik M. Roijers[a], Enda Howley[b], Patrick Mannion[b], Niel Hens[d], Ann Nowé[a] and Pieter Libin[a]

[a]Vrije Universiteit Brussel, BE
[b]National University of Ireland Galway, IE
[c]University of Antwerp, BE
[c]Hasselt University, BE

Abstract. Infectious disease outbreaks can have a disruptive impact on public health and societal processes. As decision-making in the context of epidemic mitigation is multi-dimensional hence complex, reinforcement learning in combination with complex epidemic models provides a methodology to design refined prevention strategies. Current research focuses on optimizing policies with respect to a single objective, such as the pathogen's attack rate. However, as the mitigation of epidemics involves distinct, and possibly conflicting, criteria (i.a., mortality, morbidity, economic cost, well-being), a multi-objective decision approach is warranted to obtain balanced policies. To enhance future decision-making, we propose a deep multi-objective reinforcement learning approach by building upon a state-of-the-art algorithm called Pareto Conditioned Networks (PCN) to obtain a set of solutions for distinct outcomes of the decision problem. We consider different deconfinement strategies after the first Belgian lockdown within the COVID-19 pandemic and aim to minimize both COVID-19 cases (i.e., infections and hospitalizations) and the societal burden induced by the mitigation measures. We evaluate the solution set that PCN returns, and observe that it explored the whole range of possible social restrictions, leading to high-quality trade-offs, as it captured the problem dynamics. In this work, we demonstrate that multi-objective reinforcement learning adds value to epidemiological modeling and provides essential insights to balance mitigation policies.

This work was published in the journal *Expert systems with Applications* and can be freely accessed via this link:https://doi.org/10.1016/j.eswa.2024.123686.

1 Introduction

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 [6]. 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.

In this work, 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 [13]. 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.

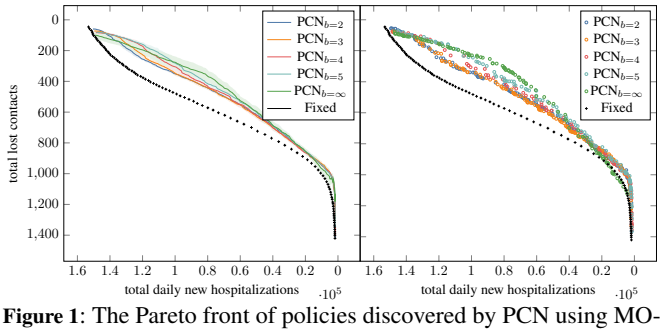

Figure 1: The Pareto front of policies discovered by PCN using MO-BelCov, showing the different compromises between the number of hospitalizations and the number of lost contacts, for set of budgets.

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

2 Methods

Stochastic compartment model 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].

Intervention strategies The social interactions of the different age-groups are modeled with a social contact matrix $\boldsymbol{C} = \boldsymbol{C}_{\text{home}} + \boldsymbol{C}_{\text{work}} + \boldsymbol{C}_{\text{transport}} + \boldsymbol{C}_{\text{school}} + \boldsymbol{C}_{\text{leisure}} + \boldsymbol{C}_{\text{other}}$, where 6 different social environments are modeled explicitly [12]. To model different types of non-pharmaceutical interventions, we consider a contact reduction function that imposes a proportional reduction of work (including transport) p_w, school p_s and leisure p_l contacts $\hat{C}(p_w, p_s, p_l) = \boldsymbol{C}_{\text{home}} + p_w(\boldsymbol{C}_{\text{work}} + \boldsymbol{C}_{\text{transport}}) + p_s\boldsymbol{C}_{\text{school}} + p_l(\boldsymbol{C}_{\text{leisure}} + \boldsymbol{C}_{\text{other}})$. At each timestep (here one week), our RL agent will modulate $p_w, p_s, p_l \in [0, 1]$ to control the spread of the epidemic.

Action budget In the context of mitigation policies, consistency is important and policies that impose changes too frequently will be hard to adhere to. As such, we introduce a *budget* regarding the number of times a policy can change over the duration of the episode. To facilitate this, we maintain a budget for each of the actions. Concretely, when the action changes, i.e., if the social restriction proposed by the policy is different from the one that is currently in place, we reduce the budget for that action by one. We only allow action changes as long as there is budget left.

Pareto Conditioned Networks (PCN) 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 PCN, a method that uses a single neural network to encompass all non-dominated policies [8] and designed for MOMDPs with discrete action-spaces to the continuous action-space setting. With this continuous action variant of PCN, we explore the Pareto front of multi-objective COVID-19 mitigation policies. 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.

3 Results

We learn a coverage set (see Fig. 1) that ranges from imposing minimal restrictions to enforcing many restrictions. As a comparison, we execute a baseline which consists of a set of 100 fixed policies, that iterate over all the possible social restriction levels. Regardless of the imposed budget, we notice that the coverage sets discovered by PCN almost completely dominate the coverage set of the baseline, demonstrating that there are better alternatives to the fixed policies. 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, PCN learns policies that drastically reduce the total number of new hospitalizations (e.g., more than 20000) for the same social burden. Moreover, we observe that the difference is concentrated around the less restrictive policies in terms of social burden. We postulate that this region contains the most complex policies, as these try to maintain as much social freedom as possible, while containing the number of hospitalisations (see Fig. 2).

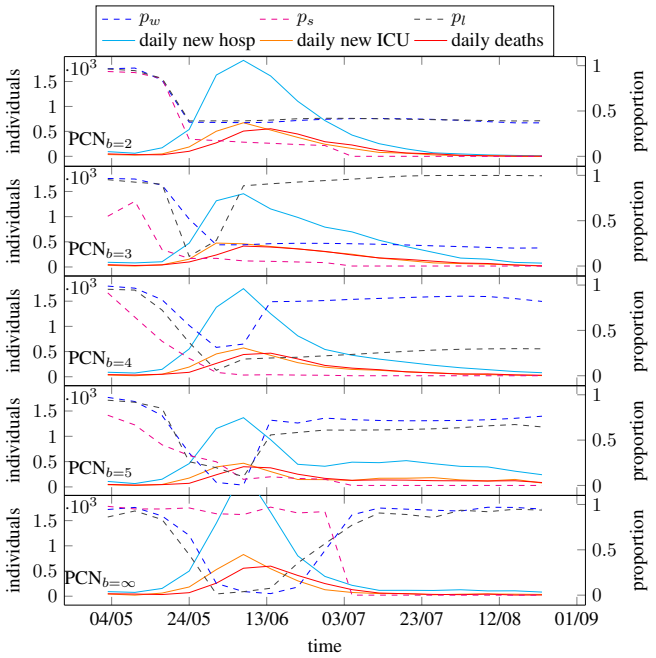

Figure 2: Execution of the policies attaining a number of hospitalisations around 80000, for different budgets. From top to bottom we display the policy executions with budget 2,3,4,5 and no-limit, respectively.

4 Conclusion

Making decisions on how to maintain epidemic situations has important ethical implications with respect to public health and societal burden. In this regard, it is crucial to approach this decision making from a balanced perspective, to which end we argue that multi-objective decision making is essential. In this work, we establish a novel approach, i.e., an expert system, to study multi-faceted policies, and this approach shows great potential to study future epidemic mitigation policies. Moreover, 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 [13], 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 [5]. We show that multi-objective reinforcement learning provides decision maker with insightful and diverse alternatives on real-world problems. PCN automatically learns all Pareto-efficient trade-offs. It explored the whole range of possible social restrictions, which led to many alternative trade-offs between these extreme policies. Furthermore, we show that action budgets can act as a regulariser that facilitates learning realistic policies that can be easily conveyed to decision makers. Finally, we notice an inflection point on the right-side of the Pareto front, indicating that taking extreme measures (which can be computed manually) may not be necessary to root out the infection while minimizing the number of hospitalisations.

Acknowledgements

C.F.H. is funded by the University of Galway Hardiman Scholarship. This research was supported by funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" program. This work also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant number 101003688 – EpiPose project). P.J.K.L. gratefully acknowledges support from FWO via postdoctoral fellowship 1242021N and the Research council of the Vrije Universiteit Brussel (OZR-VUB via grant number OZR3863BOF). N.H. acknowledges support from the Scientific Chair of Evidence-based Vaccinology under the umbrella of the Methusalem framework at the University of Antwerp. N.H. and A.N. acknowledge funding from the iBOF DESCARTES project (reference: iBOF-21-027). L.W. gratefully acknowledges support from FWO postdoctoral fellowship 1234620N. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. A.N. acknowledges the support by the FWO COVID-19 research project G0H0420N. P.L. and L.W. acknowledge support from FWO grant G059423N.

References

[1] S. Abrams, J. Wambua, E. Santermans, L. Willem, E. Kuylen, P. Coletti, P. Libin, C. Faes, O. Petrof, S. A. Herzog, 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] Z. Chen, A. S. Azman, X. Chen, J. Zou, Y. Tian, R. Sun, X. Xu, Y. Wu, W. Lu, S. Ge, et al. Global landscape of sars-cov-2 genomic surveillance and data sharing. *Nature genetics*, 54(4):499–507, 2022.

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

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

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

[6] P. J. K. Libin, A. Moonens, T. Verstraeten, F. Perez-Sanjines, N. Hens, P. Lemey, and A. Nowé. Deep reinforcement learning for large-scale epidemic control. In Y. Dong, G. Ifrim, D. Mladenić, C. Saunders, and S. Van Hoecke, editors, *ECML*, pages 155–170, Cham, 2021. Springer International Publishing. ISBN 978-3-030-67670-4.

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

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

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

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

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

[12] L. Willem, T. Van Hoang, S. Funk, P. Coletti, P. Beutels, and N. Hens. Socrates: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for covid-19. *BMC Research Notes*, 13(1): 1–8, 2020.

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