Using Reinforcement Learning for Operating Educational Campuses Safely during a Pandemic

Elizabeth Ondula and Bhaskar Krishnamachari
{ondula, bkrishna}@usc.edu

Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089

Abstract

The COVID-19 pandemic has brought a significant disruption not only on how schools operate but also affected student sentiments on learning and adoption to different learning strategies. We propose CampusPandemicPlanR, a reinforcement learning-based simulation tool that could be applied to suggest to campus operators how many students from each course to allow on a campus classroom each week. The tool aims to strike a balance between the conflicting goals of keeping students from getting infected, on one hand, and allowing more students to come into campus to allow them to benefit from in-person classes, on the other. Our preliminary results show that reinforcement learning is able to learn better policies over iterations, and that different Pareto-optimal tradeoffs between these conflicting goals could be obtained by varying the reward weight parameter.

Introduction

As a 2020 World Bank report titled "The COVID-19 pandemic: Shocks to education and policy responses " states, "The COVID-19 pandemic has already had profound impacts on education by closing schools almost everywhere in the planet, in the largest simultaneous shock to all education systems in our lifetimes" (Bank 2020) Bank. A number of schools have been temporarily closed and in many cases, adoption to new learning environments has affected student sentiments (Duong et al. 2020) and posed safety challenges as well as challenges to maintaining engagement and achieving learning outcomes (Khamees et al. 2020; Besser, Flett, and Zeigler-Hill 2020). Developing and implementing smart operational strategies under pandemic uncertainties is key to maximize overall health and safety for students while at the same time also maximizing their learning opportunities through in-person interactions wherever possible. In particular, there is a need for tools that school administrators can use to make trade-offs between learning objectives and safety of school community members. We approach this operational decision problem by designing and developing CampusPandemicPlanR, a simulation environment designed to assist in planning and scheduling of educational activities in the midst of a pandemic.

Modeling

CampusPandemicPlanR incorporates 1) a general school that includes students, teachers, courses and classrooms 2) COVID-19 transmission model. Tthe problem of finding operational strategies for campus safety is modeled as follows:

- **States**: The observed state is defined as follows. At the beginning of the nth week, it is represented as the following tuple: $\langle I_1(n), I_2(n) \ldots I_C(n), \rho_c(n) \rangle$, where $I_i(n)$ represents the percentage of infected students in the ith course in week n and the community risk level in week n is $\rho_c(n)$. For simplicity and efficiency, the observation space is discretized into a set of levels. Specifically, we use 0, 1, and 2, to represent the ranges between $0 - 33\%$, $33 - 66\%$ and $66\% - 100\%$ respectively. This could be easily modified to accommodate a more fine-grained discretization at the expense of greater storage and computational complexity for the reinforcement learning.

- **State Transition:** We use a simple, approximate model loosely based on the well-known SIR model (Cooper, Mondal, and Antonopoulos 2020) to capture how many students in a classroom become infected given the initial condition and action of how many students to allow to come into the class. The model also takes into account the community risk. It is given as follows:

$$I(n+1) = min(c_1\rho_c A_i(n)^2 + c_2 I_i(n)A_i(n), A_i(n)) \tag{1}$$

- **Actions**: The action that the agent takes is a proposal to the administrator for what percentage of students from each class to allow to be on campus. Again, for simplicity, we discretize the action also to 3 levels for each course corresponding to (attend online)$0\%, 50\%$, and (all students allowed)100% respectively.

- **Rewards**: The reward is a weighted combination of two terms. One for total number of allowed students on campus across all courses (this is a positive because it gives some educational benefit) and one for the total number of infected students (this is a negative term). The reward is calculated as follows, using a weight parameter $\alpha \in [0, 1]$

$$R(n) = \alpha \sum_{i=1}^{C} A_i(n) - (1 - \alpha) \sum_{i=1}^{C} I_i(n) \qquad (2)$$

- **Environment: Gym campus-v0**: We use OpenAI Gym (Brockman et al. 2016) (a toolkit used by many researchers for developing reinforcement learning algorithms) to develop the environment library campus-v0. The Gym API defines an interface to integrate various reinforcement learning algorithms. In our current version, we have encoded the action and observations from the environment into a finite countable state and action space.

- **Agent**: The goal is to an agent to learn by experiencing sequences of actions by estimated the value of each action any possible state. The Q-value of the action taken is the long-term expected reward under a policy. The estimated Q-values are updated after every *step* by progressively updating the difference between the current estimate and the reward obtained based on the Bellman expectation equation (Sutton, Barto et al. 1998)

Preliminary numerical results

We evaluate, the RL Agent trained with 5000 episodes. The fixed hyper-parameters and reward parameter varied after each run are summarized in the table below.

Q-learning algorithm hyper-parameters	
Hyper-parameter	Value
Learning rate	0.1
Discount factor	0.9
Exploration rate	0.2
Reward weight parameter	varied between (0.1 - 0.9)

Figure 1 shows an increasing aggregated discounted reward values over 3000 training episodes averaged over 10 runs with 95% confidence intervals. It indicates that better policies are being learned over these episodes. In figure 2 the trade-off between the number of students allowed on campus and the number that are infected, as the reward weight parameter α is varied, is shown. The RL agent was trained using different values of α. We can see the trade-off clearly – when α is high, the learned policy is such that a higher number of students are allowed and more of them get infected, and when α is low, the learned policy is such that few students are allowed on campus and thus few of them get infected.

Conclusion

The current phase of this research focused on the modeling and development of a reinforcement learning environment where decision agents can be interfaced and trained to learn operational strategies during a pandemic for different educational campus models. For the future, we would like to incorporate epidemic spread models situated for campuses and make use of real campus datasets (such as in (Hekmati et al. 2021)).
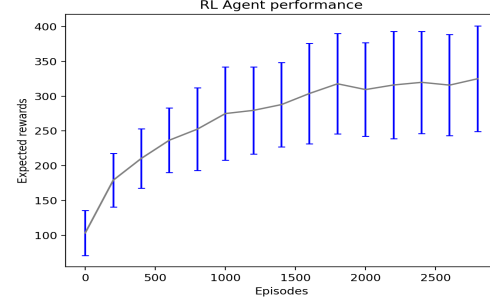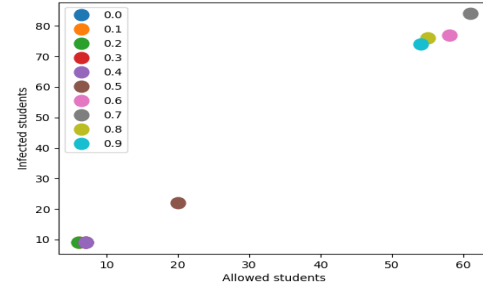

Figure 1: Episode Rewards

Figure 2: Trade-off between infected and allowed students as α is varied

References

Bank, W. 2020. The COVID-19 pandemic: Shocks to education and policy responses.

Besser, A.; Flett, G. L.; and Zeigler-Hill, V. 2020. Adaptability to a sudden transition to online learning during the COVID-19 pandemic: Understanding the challenges for students. *Scholarship of Teaching and Learning in Psychology*.

Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; and Zaremba, W. 2016. Openai gym. *arXiv preprint arXiv:1606.01540*.

Cooper, I.; Mondal, A.; and Antonopoulos, C. G. 2020. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139: 110057.

Duong, V.; Pham, P.; Yang, T.; Wang, Y.; and Luo, J. 2020. The ivory tower lost: How college students respond differently than the general public to the COVID-19 pandemic. *arXiv preprint arXiv:2004.09968*.

Hekmati, A.; Luhar, M.; Krishnamachari, B.; and Matarić, M. 2021. Simulation-Based Analysis of COVID-19 Spread Through Classroom Transmission on a University Campus. *arXiv preprint arXiv:2104.04129*.

Khamees, D.; Brown, C. A.; Arribas, M.; Murphey, A. C.; Haas, M. R.; and House, J. B. 2020. In crisis: medical students in the COVID-19 pandemic. *AEM Education and Training*, 4(3): 284–290.

Sutton, R. S.; Barto, A. G.; et al. 1998. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge.