Using reinforcement learning to determine when to provide human support in quitting smoking with a virtual coach

Author
Shirley Li
s.x.li@student.tudelft.nl

Supervisor
Nele Albers

Responsible Professor
Willem-Paul Brinkman

1 Introduction

- Smoking kills approximately 8 million people per year [1]
- eHealth applications with virtual coaches have emerged to assist in quitting smoking
- Studies have shown that additional human support is beneficial [2][3]
- Due to the budgetary constraints and limited availability of human coaches, it is important to be able to decide when to provide human feedback to optimize effectiveness
- This study explores using reinforcement learning to determine this

2 Research Question

How effective is a reinforcement learning model in determining when to provide human feedback that optimizes the effort people spend on their activities and the chance that they stay in the intervention?

TU Delft

3 Methodology

Train and evaluate reinforcement learning model with data collected from 678 smokers/vapers [4]

Reinforcement learning model

- **User states**
  - 2 features: appreciation of human support and self-efficacy
  - 0 denotes a relatively low score, 1 medium, and 2 high.
- **Actions**: providing human feedback (1) or not (0)
- **Reward**
  - Represents users' behavior: effort spent on activities and likelihood of returning to the intervention
  - Map effort and return responses to [-1, 1] with 0 as mean and use weighted sum of both rewards
  - Deduct a cost of 0.21 if human feedback is provided
- **Discount factor**: 0.85 (favor rewards in near future versus distant future)

- Learn optimal policy π* that maximizes the expected cumulative reward over time

4 Results

State	00	01	02	10	11	12	20	21	22
π*	0	0	0	0	0	0	1	0	0

Table 1: Optimal policy for each state

Effect of optimal policy on users' states

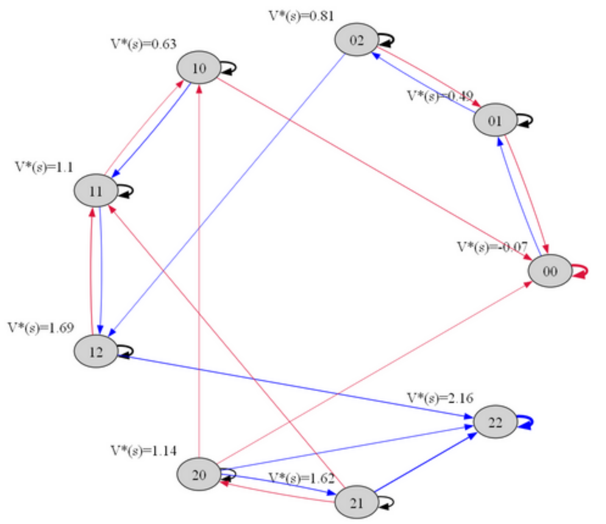

Figure 1: Transition probabilities under π*, only probabilities of at least 1/|S| are shown and a thicker line denotes a higher probability

Receiving human support in state 20 tends to move people to states 21 and 22, improving their self-efficacy and expected behavior

Overall, people tend to move to states with a higher expected reward (blue lines)

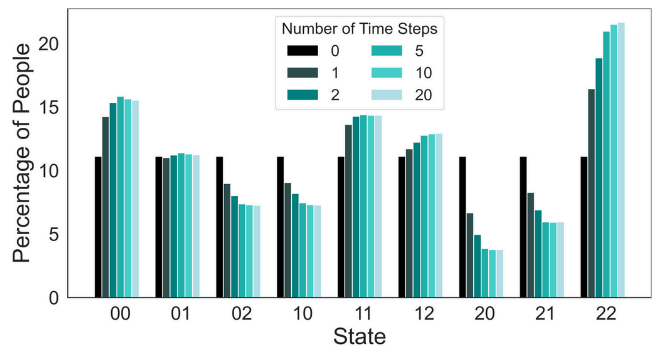

Figure 2: Percentage of people in each state after following the optimal policy for various numbers of time steps

After following the optimal policies for multiple steps, most people get to the better states (11, 12, 22), where the expected reward is higher

However, people also remain in state 00 where the expected reward is lowest

Effect of optimal vs. sub-optimal policy on users' behavior

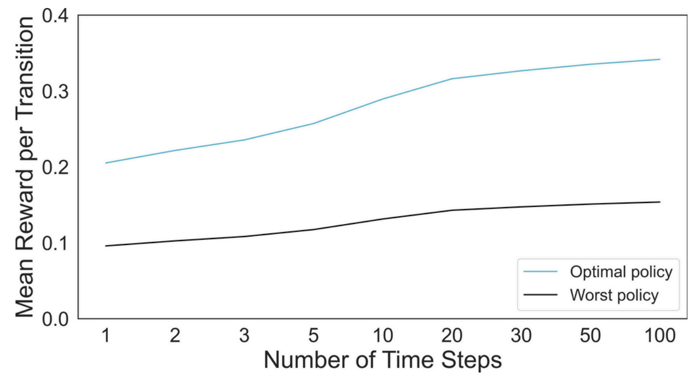

Figure 3: Mean reward per transition over time while following the optimal policy and the worst policy.

Users' behavior improves more following optimal policy compared to worst policy

Effect of different cost factors in reward function

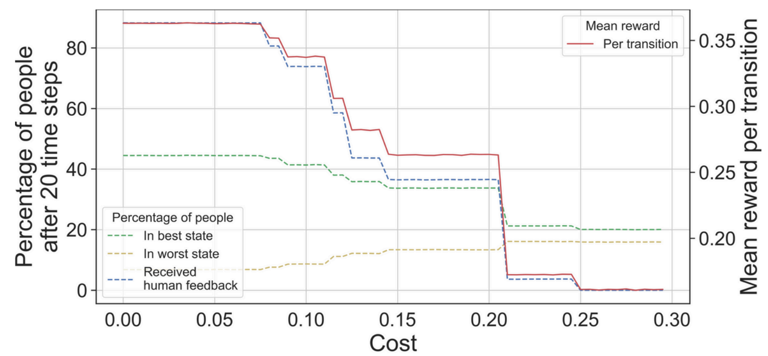

Figure 4: Simulation results after 20 time steps for different cost factors.

A higher cost factor seems to be more cost-effective as relatively less human support is required to improve users' behavior

5 Conclusion

- Reinforcement learning is effective in determining when to provide human support and increases users' effort and return likelihood
- A higher cost factor for involving a human coach in our reward function leads to more cost-effective results

6 Future work

- Run a trial to assess effectiveness in real-life scenarios
- Different reward function
  - Dynamic reward function with changing cost factor based on the availability of human coaches
  - Investigate the importance of the two objectives for smoking cessation -> possibly different weights
  - Consider other ethical principles
  - Explore more complex multi-objective reinforcement learning algorithms

References

[1] World Health Organization. "Tobacco." (2023), [Online]. Available: https://www.trimbos.nl/kennis/cijfers/roken/.
[2]P. Chikersal, D. Belgrave, G. Doherty, et al., "Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI USA: ACM, Apr. 2020, pp. 1–16. [Online]. Available: https : / / dl . acm . org / doi / 10 . 1145 / 3313831 . 3376341
[3] Y.-C. Lee, N. Yamashita, and Y. Huang, "Exploring the Effects of Incorporating Human Experts to Deliver Journaling Guidance through a Chatbot," in Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW1, pp. 1–27, Apr. 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3449196
[4] N. Albers and W.-P. Brinkman, "Perfect fit - learning when to involve a human coach in an ehealth application for preparing for quitting smoking or vaping," 2024. DOI: https://doi.org/10.17605/OSF.IO/78CNR. [Online]. Available: https://osf.io/78cnr.