Integrating Human Feedback in a Virtual Smoking Cessation Coach: Optimizing Behavioral and Identity Outcomes with Reinforcement Learning.

Ghiyath Alaswad
Responsible Professor: Dr. ir. Willem-Paul Brinkman,
Supervisor: Ir. Nele Albers
EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
August 18, 2024

Name of the student: Ghiyath Alaswad
Final project course: CSE3000 Research Project
Thesis committee: Willem-Paul Brinkman, Nele Albers, Gosia Migut

Abstract

The use of chatbots in eHealth applications has grown significantly, providing an interactive platform for users to improve their health and lifestyle. To enhance these applications, human feedback has been incorporated, with a coach offering personalized guidance to help users achieve their goals. However, deciding whether to provide human feedback or rely solely on automated systems remains a challenge, especially given the cost constraints associated with human involvement. This study presents a reinforcement learning (RL) model designed to optimize this decision by analyzing users' states and predicting the potential benefit of such interventions. The model was trained on data from a longitudinal study involving over 500 daily smokers and vapers who interacted with a virtual coach across multiple sessions. Our findings indicate that the RL model effectively assists in determining whether human feedback is valuable, increasing the likelihood of behavior change, enhancing users' quitter identities, and further reducing smoking/vaping frequency. This research contributes to the development of more effective, resource-efficient eHealth interventions.

1 Introduction

According to the World Health Organization (WHO), there are 1.3 billion tobacco users worldwide [1]. Tobacco claims more than eight million lives each year, and generally, it is responsible for the death of about half of its users who do not quit [1]. Therefore, it is considered a significant public health threat [1].
The same resource (WHO) indicates that most smokers would like to quit smoking. While counseling and medication can double a smoker's chance of quitting, assistance is only available in 32 countries based on current resources [1]. This represents only about 16% of the global population. However, due to a lack of available resources, even in well-resourced countries like the Netherlands [13] it is not possible to give personal help to all people who need it.
One suggested solution is e-health applications, which are attracting more researchers and organizations nowadays and are considered a "promising way to support people in changing their behavior" [3, 4]. The reason behind this could be the widespread use of smartphones, which are a part of everyone's life and make such applications reachable for the user anytime so that they can track their progress easily, and helps organizations reach different segments of users with the eHealth applications. A commonly used strategy in e-health applications is coaching people to achieve different goals, such as changing bad habits or behaviors or improving their lifestyle, etc. While assisting those people to reach their goals by a human expert would be costly, an effective solution to solve this is to interact with users using chatbots, also called conversational agents [5, 6]. Especially after the rapid development of artificial intelligence, chatbots based on AI technology can interact with people in a more human-like way and have natural conversations [7]. Given the widespread issue of smoking, it has been targeted to be addressed using these applications [6].
These systems offer a cost-effective alternative to traditional human coaching, which can be resource-intensive and limited in availability. However, fully automated interventions often face limitations in their effectiveness, particularly when dealing with the complexities of individual behavior change.

One potential solution to this problem is the integration of human feedback into eHealth applications. By providing personalized guidance from a human coach, these applications can potentially enhance their effectiveness as shown by Chikersal et al.[16], and Lee et

al.[17]. However, a critical challenge remains: deciding whether human feedback is necessary for each user interaction, given the costs and limited availability of human resources. The effectiveness of such feedback likely depends on the user's current state, such as their motivation, mood, and readiness to quit smoking.

To address this challenge, this study explores the use of reinforcement learning (RL) to optimize the decision-making process in providing human feedback within a virtual smoking cessation coach. Previous research has shown that RL can be effectively applied to personalize the predictions based on the user's current/future states [8, 9].
This study aims to develop an RL model that analyzes users' states and predicts whether human feedback is likely to be beneficial, thereby improving both behavioral outcomes especially smoking/vaping frequency in this case, and quitter self-identity, which plays an important role in helping users reach their goals as shown by Meijer et al[18].

Through a longitudinal study involving over 500 daily smokers and vapers, this research collected extensive data on user interactions with a virtual coach. The RL model was trained on this data to learn whether human feedback should be provided to maximize the effectiveness of the intervention. By balancing the benefits of human feedback with the associated costs, this approach offers a path toward more efficient and impactful eHealth solutions.

2 Methodology

As mentioned in the introduction, this research consists of two parts, the first part is about designing and training a reinforcement model that decides whether to provide feedback for the user based on his state, and the second part is about analyzing the effectiveness and reliability of the designed model. The reinforcement learning includes several parts, namely:

- Environment.
- Agent which is the virtual coach in this research.
- Data which are used to train the model.
- Reinforcement learning model.
- Policy is the strategy the agent uses to determine the action to be taken.

Let's elaborate more on those parts for our research:

2.1 Agent: Virtual coach

The virtual coach, named Kai [10], was developed to interact with participants during five structured sessions aimed at helping them prepare to quit smoking. During each session, Kai asked participants about their current states (e.g., mood, energy level) and suggested specific activities based on their responses. In some instances, participants also received feedback from a human coach, though this was limited to 20% of interactions due to budget constraints. This selective feedback was designed to provide more tailored support and was based on the participants' earlier responses.

2.2 Data

The study initially involved over 500 daily smokers and vapers, each of whom was required to complete pre-screening and post-session questionnaires as well. These questionnaires captured a range of data, including demographic information (age, gender, education level), smoking/vaping frequency, and self-reported quitter identity. Participants were also asked about their perceptions of human support and their motivation to quit smoking. For this research, we only considered the data of the 459 users who completed all sessions, including both the pre-and post-questionnaires. The study and the data are available online at OSF-form[19]

In developing the model, we focused on users' responses to Kai's questions about their current state during each session. To assess changes in smoking/vaping frequency, we compared the values reported in the pre-questionnaires to those in the post-questionnaires. Additionally, we measured identity change using three items based on Meijer et al.[14], comparing the identity scores from the pre-questionnaires to those from the post-questionnaires to evaluate the impact of the intervention.

2.3 Data Preprocessing

The data preprocessing involved several critical steps to prepare the dataset for analysis and model training:

- **Filtering and Cleaning**: Data was loaded from multiple sources, including pre-screening questionnaires, session data, and post-session questionnaires. Only users who completed all sessions, including the pre- and post-questionnaires, were included in the final dataset, resulting in 459 participants.
- **Feature Selection and Transformation**: Irrelevant columns and rows were excluded to focus on the most relevant data for the study. Session data was transformed by pivoting response types into columns, and categorical responses were mapped to numerical values for consistency.
- **Combining Datasets**: The pre-screening, session, and post-session data were merged into a single dataset using a common identifier. Similar columns, such as smoking and vaping frequencies, were combined to reduce redundancy and simplify the analysis.
- **Reward Calculation**: Rewards were calculated based on changes in smoking/vaping frequency and identity scores from pre- to post-intervention. These rewards were added as new columns in the dataset to be used in the reinforcement learning model.
- **Creating RL Samples**: Using the preprocessed data, RL samples were created, resulting in 2,294 samples. These samples include the user states, actions (whether human support was provided), and the corresponding rewards, which are essential for training the RL model.

2.4 Reinforcement learning model

To accomplish the first part of the research, a reinforcement model is designed as a Markov Decision Process (MDP) which is defined as a tuple $<S, A, R, T, \gamma>$ where:

- S is the state space,

- A is the action space,
- R is the reward function,
- T defines the transitions between states,
- γ is the discount factor.

An agent(virtual coach) operating within a Markov Decision Process (MDP) aims to learn the best action to take in each possible state. The goal is to maximize the expected sum of rewards over time, taking into account that future rewards are often less valuable than immediate ones. This process is formalized by the equation $E\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$, where t represents time steps, r_t is the reward at time step t, and γ is the discount factor, indicating how much future rewards affect current decisions.

In simpler terms, each time the agent takes an action in a particular state, it moves to a new state and receives a reward. The probability of moving to a specific new state, given the current state and action, is denoted by $p(s'|s, a)$. The agent's task is to figure out which actions to take in each state to maximize its long-term rewards.

States:
The set of states that will be used to train the model is the user's current states which are:

- state energy: is based on the question "How much energy do you have?" , rated on a scale from 0 ("none") to 10 ("extremely much").
- Appreciation of human feedback based on the question "How would you view receiving a feedback message from a human coach after this session?", rated on a scale from -10 ("very negatively") to 10 ("very positively"), with 0 labeled as "neutral."
- state importance: is based on the question "How important is it to you to prepare for quitting smoking now?", rated on a scale from 0 ("not at all important") to 10 ("desperately important") based on the paper by Rajani et al.[15].
- state Self-efficacy for preparing for quitting smoking based on the question "How confident are you that you can prepare for quitting smoking now?", rated on a scale from 0 ("not at all confident") to 10 ("highly confident"). The scale is adapted from McAuley's Exercise Self-Efficacy Scale.

Action:
The action space consisted of two possible actions:

- Provide human feedback.
- Do not provide human feedback.

Reward Function:
To design the reward function, several aspects were considered. The model focuses on changing two behaviors: smoking/vaping frequency and changes in the user's identity. Both behaviors were incorporated into the reward function by comparing the user's smoking/vaping frequency and identity scores from the pre-screening questionnaire to those in the post-questionnaire. This comparison accounts for both desirable and undesirable changes.

Given this focus, we implemented a sparse reward function. In this approach, rewards were set to zero during the sessions and only assigned after all sessions were completed, based on the values from the post-questionnaire.

The combined sparse reward function was designed as follows:

- Smoking/vaping frequency (50%)
- Change in user identity score (50%)

To address budget constraints, a cost factor was added as a penalty to the reward function, resulting in the final structure:

- Smoking/vaping frequency (50%)
- Change in user identity score (50%)
- Cost factor (-0.01)

We explored different approaches to designing the reward function:

- **Behavioral Weights**: We experimented with several weight distributions for different behaviors and found that giving equal weight (50/50) to each behavior worked best.
- **Cost Factor Values**: We tested various values for the cost factor to evaluate their impact on the outcomes.
- **Improvement Rewards**: We compared two methods for rewarding improvements: one that assigned rewards based on the percentage of improvement and another that gave a fixed reward for reaching a goal.

Discount factor:
The discount factor is set at 0.85, reflecting a value close to one, which emphasizes a greater focus on long-term changes following the approach of Albers et al.[9]

3 Results and Discussion

In this section, we present the results, discuss various approaches, and provide the reasoning behind each decision made during the research. We will also address the sub-research questions:

3.1 How can the state space be reduced to improve the accuracy of the results, which states to consider as part of the model, and how to choose these states?

To tackle this question, we explored several methods for reducing the state space. We first looked into the G-algorithm proposed by Chapman and Kaelbling [12], which was inspired by the approach used by Albers et al. [9]. However, implementing this algorithm proved difficult due to the nature of our model's sparse reward function, where rewards are only given at the very end, after the last session and transition. This made it challenging to achieve clear and actionable results.

We also experimented with using mutual information to reduce the state space. Mutual information is a method used to evaluate how much the input features (or states) are dependent on the target variable (or actions in reinforcement learning). It helps in selecting the most important features that contribute to predicting the target, as highlighted by Covert et al.[20].

In the end, given the complexities and challenges of effectively applying these algorithms, we decided to stick with the four available states in the model. This approach allowed us to maintain simplicity and effectiveness, avoiding unnecessary complications in our analysis.

3.2 What is the optimal reward function, considering the cost of human feedback and the limitations of availability?

As mentioned previously, we utilized a sparse reward function in our model. This approach assigns rewards only at the end of the sessions, meaning all intermediate steps yield a reward of zero. This increases the difficulty of learning since the agent does not receive immediate feedback on its actions, making it harder to discern whether a particular action was beneficial.

The reward calculation focuses on two key behaviors: smoking/vaping frequency and identity change. These behaviors are assessed using data collected during the pre- and post-questionnaires. For smoking/vaping frequency, we compared the user's response from the pre-questionnaire to the post-questionnaire, calculated the difference, and then normalized the result to obtain a value between -1 and 1.

For identity change, we measured it using three items based on Meijer et al. [14]. The final identity score comprises nine values: three related to smoker/vaper identity, three related to quitter identity, and three related to non-smoker/vaper identity. We compared these values from the pre-questionnaire to the post-questionnaire, considering both desirable and undesirable changes. An increase in the smoker/vaper identity score was considered undesirable, while an increase in the non-smoker/vaper and quitter identity scores was considered desirable. We calculated the difference and normalized the results to get a value between -1 and 1.

After obtaining the individual rewards, we computed the weighted reward as mentioned earlier. Additionally, we included a negative value as a cost factor whenever human feedback was provided.

3.3 How effective is the model, and how reliable are the predictions it proposes compared to other algorithms?

4 Responsible Research

4.1 Ethical Considerations

This research, aimed at developing behavioral interventions to assist individuals in reducing or quitting smoking and vaping, was conducted with a strong emphasis on ethical considerations, particularly regarding participant privacy and the careful delivery of feedback interventions. All data was fully anonymized to ensure that no personal identifiers could be linked to any participant. The model was designed to provide human feedback only when deemed genuinely beneficial, minimizing unnecessary or potentially distressing interventions. By using a sparse reward function, feedback was strategically provided at the end of sessions, ensuring a balance between effective support and the learning process.

The study by Albers et al.[19] from which the data was gathered received approval from the Human Research Ethics Committee of the Delft University of Technology (Letter of Approval number: 3683), reflecting our commitment to maintaining high ethical standards in research involving human participants.

4.2 Reproducibility of Methods

Ensuring the reproducibility of our research was a central focus. Each step of the data preprocessing, model training, and analysis was thoroughly documented, allowing other researchers to replicate our findings. The datasets used in this study will be published on OSF, and the associated scripts and code will be made available on 4TU.ResearchData[21].

In conclusion, this research was conducted with a strong emphasis on ethical responsibility and reproducibility, contributing to the broader goal of transparent and accountable scientific inquiry in the field of behavioral interventions.

5 Conclusions and Future Work

This research explored the application of reinforcement learning to support behavioral changes, specifically targeting smoking and vaping cessation. The primary goal was to determine when human feedback should be provided to maximize positive behavioral outcomes, focusing on key behaviors: smoking/vaping frequency and identity change.

The model utilized a sparse reward function, where rewards were assigned only at the end of the sessions. This approach increased the challenge for the model, as it had to make decisions without immediate feedback, making it difficult to evaluate the effectiveness of individual actions throughout the process.

We experimented with various methods to reduce the state space, including the G-algorithm and mutual information techniques. However, due to the complexities introduced by the sparse reward structure and the inherent challenges of these algorithms, we ultimately decided to use the four original states in the model. This decision maintained a balance between simplicity and the modelâs ability to effectively predict outcomes.

The reward function was carefully designed to capture changes in smoking/vaping frequency and identity. By normalizing these changes to values between -1 and 1, the reward system accurately reflected the desired outcomes. Additionally, a cost factor was introduced to penalize unnecessary human feedback, encouraging the model to optimize its decision-making process regarding whether or not to provide human intervention.

The study's findings highlight the importance of a well-structured reward function, particularly in scenarios where feedback is sparse and delayed. The model demonstrated an ability to make informed decisions about the necessity of human feedback based on changes in the key behaviors of smoking/vaping frequency and identity.

5.1 Future Work

While the current study provides a solid foundation for using reinforcement learning to guide behavioral interventions, several areas warrant further exploration. One of the key challenges encountered was the high dimensionality of the state space. Although the model was able to operate effectively with the selected states, future research should focus on developing more robust methods for state space reduction.

Investigating advanced techniques for dimensionality reduction, such as deep learning-based feature extraction or state aggregation methods, could significantly enhance the model's efficiency and accuracy. Additionally, integrating more dynamic reward structures that provide feedback at multiple stages, rather than only at the conclusion of the sessions, may offer a more nuanced understanding of the intervention's impact and improve the learning process.

Furthermore, expanding the model to incorporate additional behavioral factors beyond smoking/vaping frequency and identity change could provide a more comprehensive framework for supporting behavioral change.

In conclusion, this study demonstrates that reinforcement learning can be effectively applied to support decisions on whether to provide human feedback in behavioral change interventions. By focusing on smoking/vaping frequency and identity change, and through careful reward design, the model provides a promising approach to enhancing personalized support strategies in smoking and vaping cessation. Future research should explore more dynamic reward structures and consider additional behavioral factors to further refine and improve intervention strategies.

6 Acknowledgments

I would like to express my sincere gratitude to Nele Albers and Willem-Paul Brinkman for their invaluable guidance and support throughout this project. Their expertise and insights were instrumental in shaping the direction of this research. I am also deeply appreciative of the foundational code they provided, which served as a critical basis for the development of the model used in this study. Their contributions have been essential to the success of this work.

This work is part of the multidisciplinary research project Perfect Fit, which is supported by several funders organized by the Netherlands Organization for Scientific Research (NWO), program Commit2Data - Big Data & Health (project number 628.011.211). Besides NWO, the funders include the Netherlands Organisation for Health Research and Development (ZonMw), Hartstichting, the Ministry of Health, Welfare, and Sport (VWS), Health Holland, and the Netherlands eScience Center.

7 Refrences:

[1] https://www.who.int [Accessed: May.07.2024].

[2] `https://perfectfit-research.com/en`**.**

[3] Nele Albers, Mark A. Neerincx, and Willem-Paul Brinkman. 2023. Persuading to Prepare for Quitting Smoking with a Virtual Coach: Using States and User Characteristics to Predict Behavior.

[4] Teresa Amabile and Steven Kramer. 2011. The progress principle: Using small wins to ignite joy, engagement, and creativity at work. Harvard Business Review Press, Boston, USA.

[5] Liliana Laranjo, Adam G Dunn, et al. Conversational agents in healthcare: a systematic review, Journal of the American Medical Informatics Association, Volume 25, Issue 9, September 2018, Pages 1248â1258, https://doi.org/10.1093/jamia/ocy072

[6] Pereira, J., DÃaz, Ã. Using Health Chatbots for Behavior Change: A Mapping Study. J Med Syst 43, 135 (2019). https://doi.org/10.1007/s10916-019-1237-1

[7] Zhang J, Oh Y, Lange P, Yu Z, Fukuoka Y Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint, J Med Internet Res 2020;22(9):e22845,https://www.jmir.org/2020/9/e22845, DOI: 10.2196/22845

[8] den Hengst, Floris et al. âReinforcement Learning for Personalization: A Systematic Literature Reviewâ. 1 Jan. 2020 : 107 â 147.

[9] Albers N, Neerincx MA, Brinkman W-P (2022) Addressing peopleâs current and future states in a reinforcement learning algorithm for persuading to quit smoking and to be physically active. PLoS ONE 17(12): e0277295. https://doi.org/10.1371/journal.pone.0277295

[10] Nele Albers. (2024). PerfectFit-project/virtual_coach_human_inv: Virtual Coach Kai for Preparing for Quitting Smoking with Human Support (v1.0). Zenodo. https://doi.org/10.5281/zenodo.11102861

[11] Nele Albers and Willem-Paul Brinkman, âPerfect fit-earning, when to involve a human coach in an ehealth application for preparing for quitting smoking or vaping,â2024. DOI: https://doi.org/10.17605/OSF.IO/78CNR.

[12] David Chapman and Leslie Pack Kaelbling. âInput Generalization in Delayed Reinforcement Learning: An Algorithm and Performance Comparisons.â In: Ijcai. Vol. 91. 1991, pp. 726â731.

[13] Een gezond vooruitzicht Synthese| Volksgezondheid Toekomst Verkenning, RIVM, 2018

[14] Meijer E, Gebhardt WA, Van Laar C, Kawous R, Beijk SC. Socio economic status in relation to smoking: The role of (expected and desired) social support and quitter identity. Social Science / Medicine. 2016; 162:41â49. https://doi.org/10.1016/j.socscimed.2016.06.022 PMID: 27328056

[15] N. B. Rajani, N. Mastellos, and F. T. Filippidis, âSelfefficacy and motivation to quit of smokers seeking to quit: Quantitative assessment of smoking cessation mobile apps,â JMIR Mhealth Uhealth, vol. 9, no. 4, e25030, Apr. 2021, ISSN: 2291-5222. DOI: https://doi.org/10.2196/25030. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/33929336.

[16] P. Chikersal, D. Belgrave, G. Doherty, et al., âUnderstanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention,â en, in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI USA: ACM, Apr. 2020, pp. 1â16, ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376341. Available: https://dl.acm.org/doi/10.1145/3313831.3376341.

[17] Y.-C. Lee, N. Yamashita, and Y. Huang, âExploring the Effects of Incorporating Human Experts to Deliver Journaling Guidance through a Chatbot,â en, Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW1, pp. 127, Apr. 2021, ISSN: 2573-0142. DOI: 10.1145/3449196. Available: https://dl.acm.org/doi/10.1145/3449196

[18] Meijer E, Gebhardt WA, van Laar C, van den Putte B, Evers AW. Strengthening quitter self-identity: An experimental study. Psychology health. 2018; 33(10):1229â1250. https://doi.org10.108008870446.2018.1478976 PMID: 29886765

[19] N. Albers and W.-P. Brinkman, âPerfect fit - learning when to involve a human coach in an ehealth application for preparing for quitting smoking or vaping,â 2024. DOI: https://doi.org/10.17605/OSF.IO/78CNR. Available: https://osf.io/78cnr.

[20] Covert, I.C., Qiu, W., Lu, M., Kim, N.Y., White, N.J. amp; Lee, S.. (2023). Learning to Maximize Mutual Information for Dynamic Feature Selection. Proceedings of the 40th International Conference on Machine Learning, in Proceedings of Machine Learning Research 202:6424-6447 Available from https://proceedings.mlr.press/v202/covert23a.html.

[21] G. Alaswad, Analysis code for bachelor thesis: Integrating Human Feedback in a Virtual Smoking Cessation Coach: Optimizing Behavioral and Identity Outcomes with Reinforcement Learning. August. 2024. DOI: 10.4121/0586815e-3655-4a15-b09a-2c79e8faa259