Using Reinforcement Learning to Determine When to Provide Human Support in Quitting Smoking with a Virtual Coach

This file provides information on how to reproduce our analyses for the paper "Using Reinforcement Learning to Determine When to Provide Human Support in Quitting Smoking with a Virtual Coach" by Shirley Li.

Author: Shirley Li **Date**: June 2024

Requirements

The data that we used in our study is not published at the time of writing. However, it will be published and linked to the OSF Form, which can be found here: https://osf.io/78cnr. Download the data and place the following files in the `Data` folder:

- `prolific_profile_bsc_anonym.csv`
- `sessionsdata_bsc_anonym.csv`

To be able to run the analysis code, please ensure that you have Python and Jupyter Notebooks installed.

For the analysis code for Q3 `Analysis/analysis_q3.ipynb`, you should also have Graphviz installed to be able to recreate the transition graph.

Data

The `Data` folder should contain the data that we used for the analysis questions:

- `prolific_profile_bsc_anonym.csv`
  - This contains data on the characteristics (age, gender, education level) of the participants.
- `sessionsdata_bsc_anonym.csv`
  - This contains data that we use to create the transition samples for our reinforcement learning model.

All data has been anonymized and there is an identifier `rand_id` that is used to distinguish between people.

Preprocessing

To preprocess the data and to extract the transition samples, the `preprocessing.py` file is used. This file generates the two files `all_states.csv` and `rl_samples.csv` in the `Analysis/Data` folder, which are used for the analysis questions.

Steps to Reproduce Analyses

1. Make sure you have Python and Jupyter Notebook installed and create a virtual environment to work in. Now run `pip install -r requirements.txt` to install the required libraries.
2. If desired, you can run the `preprocessing.py` file to recreate the transition samples. The `Analysis/Data` folder now already contains a version of those files that we have computed.

3. Navigate to the `Analysis` folder. Here you can find notebooks for each of the 6 research questions, as well as the feature analysis to determine our feature representation. In addition, there is the `participant_characteristics.ipynb` file to reproduce the table with participant characteristics in the Appendix.

Explanation of files and folders

This repository contains the following folders and files:

- `Analysis`: Folder containing scripts and Jupyter notebooks for our research questions.
  - `analysis_features.ipynb`: Analysis of features to determine the feature representation.
  - `analysis_q1.ipynb` to `analysis_q5.ipynb`: Notebooks for analysis of our research questions (e.g. `analysis_q1.ipynb` corresponds to Q1).
  - `Compute_RL_Actions.py`: Code for computing mean reward per action.
  - `Compute_RL_QValues.py`: Code for computing Q-values and dynamics of our RL problem.
  - `feature_selection.py`: Code for performing feature selection.
  - `participant_characteristics.ipynb`: Notebook for analysis of participant characteristics.
  - `reward_utils.py`: Utility functions for reward computation.
  - `Figures`: Folder with figures generated by the analysis notebooks.
  - `Data`: Folder containing files with transition samples created from preprocessing.
    - `all_states.csv`: Data on all states (even from people who have only been to one session).
    - `rl_samples.csv`: Data on valid transition samples of the form <s, a, r, s'>.
- `Data`: Folder with raw data files and explanation.
- `preprocessing.py`: Script for preprocessing raw data.
- `README.md`: This file.
- `requirements.txt`: File containing the required Python libraries.

Acknowledgements

Special thanks to Albers et al. who provided the analysis code that a large part of our code is adapted from. Their code was for a similar study titled "Persuading to Prepare for Quitting Smoking with a Virtual Coach: Using States and User Characteristics to Predict Behavior" and is publicly available here.