Data overview, RQ2, and RQ3

Author: Nele Albers **Date**: January 2025

This file is meant to guide you through reproducing the information we provide on our collected data as well as our analyses for RQ2 and RQ3.

Authored by Nele Albers, Francisco S. Melo, Mark A. Neerincx, Olya Kudina, and Willem-Paul Brinkman.

Types of analyses

Data overview and participant characteristics

Refer to the file "dataoverview_participantcharacteristics.ipynb" to reproduce:

- The mean effort per preparatory activity (Table C1 in the Appendix),
- Mean effort per action and combination of selected state feature values (Figure C5 in the Appendix),
- Number of samples per action and combination of selected state feature values (Figure C6 in the Appendix),
- Mean return likelihood rating per session (Results-section), and
- The participant characteristics (Table C5 in the Appendix).

Refer to the file "human_feedback_overview.ipynb" to reproduce the following numbers reported in the "Results"-section:

- The number of people noticing human feedback according to the post-questionnaire,
- The number of people reading human feedback messages according to the post-questionnaire,
- The number of people clicking on the reading confirmation links in the feedback messages, and
- The number of people with data from the next session after receiving vs. not receiving human feedback.

Refer to the file "user_weights_for_allocation_principles.ipynb" to reproduce the following:

- The weights participants assigned to the allocation principles, grouped based on the auxiliary rewards we have, from Table 6.4,
- The weights participants assigned to the allocation principles from Table C4 in the Appendix, and
- The relative weights for allocation principles corresponding to our auxiliary rewards compared to prognosis, used for our analysis for RQ3.

The files "dataoverview_participantcharacteristics.html," "human_feedback_overview.html," and "user_weights_for_allocation_principles.html" show the corresponding results as computed by us.

Feature selection for RL model

Refer to the file "feature_selection_from_nonabstracted_states.ipynb" to reproduce our selection of three base state features. The file "feature_selection_from_nonabstracted_states.html" shows the results as computed by us.

Analysis for RQ2

Refer to the file "rq2_longterm_effect_unlimited_feedback.ipynb" to reproduce our analysis on the long-term effect of unlimited human feedback. This includes:

- Figure 6.2, and
- The transition functions for giving and not giving human feedback shown in Figure C1 in the Appendix.

Refer to the file "rq2_longterm_effect_limited_feedback.ipynb" to reproduce our analysis on the long-term effect of limited human feedback. This includes:

- Policies for different human feedback costs from Table 6.2,
- Figure 6.3,
- The distribution of people across the 12 base states we observed in the first session of our longitudinal study from Figure C7 in the Appendix, and
- The mean reward and percentage of people receiving human feedback when repeating the analysis for our hypothetical live application described for RQ3, shown in Figure C8 in the Appendix.

The corresponding .html-files again show the results as computed by us.

To re-compute the dynamics and Q-values underlying the computations, refer to the file "compute_policies.py."

Analysis for RQ3

Refer to the file "rq3.ipynb" to reproduce our analysis for RQ3. This includes:

- The average percentage of people with negative return likelihood ratings in our study, used as basis for the dropout in our hypothetical live application, and
- Figure 6.4.

The corresponding .html-file again shows the results as computed by us.

Policies for return likelihood as reward

Refer to the file "policies_return_likelihood.ipynb" to reproduce:

- The optimal policies for different human feedback costs when using the return likelihood as basis for the reward (Table C2 in the Appendix).

The corresponding .html-file again shows the results as computed by us.

To re-compute the dynamics and Q-values underlying the computations, refer to the file "compute_policies_returnlikelihood_effortfeatures.py."

Steps to reproduce analyses

The reproduction of our code is based on Docker and Jupyter Notebook. Take the following steps:

1. Make sure that you have Docker installed. You can check whether you do by running `docker -v`.
2. Now choose from the following two options:
    - In the directory of this README-file, build the Docker image via `docker build . -t gbna4/humaninv2024_python`.
    - Pull the Docker image from Dockerhub via `docker pull gbna4/humaninv2024_python`.

3. Run the Docker container via `docker run -p 8888:8888 -e JUPYTER_ENABLE_LAB=yes -v <this_working_directory>:/home/jovyan/work gbna4/humaninv2024_python`, where `<this_working_directory>` is the path to the directory that this README-file is in.
4. Go to one of the links presented in the terminal upon running the Docker container to access Jupyter Notebook.
5. Open the "work"-folder in Jupyter Notebook.
6. Open one of the notebooks to reproduce the corresponding analyses.

Explanation of files and folders

This directory contains the following files and folders:

- Data: Data needed for our analyses
    - all_abstract_states_with_session.csv: All abstract states, including those of people with no state data in session 2, with corresponding session.
    - all_states: All non-abstracted states, including those of people with no state data in session 2.
    - data_rl_samples.csv: Non-abstracted transition samples.
    - data_rl_samples_abstracted[0, 1, 2][3, 2, 2].csv: Abstracted transition samples.
    - ethical_principle_relative_weights_with_prognosis: Weights for the allocation principles corresponding to our auxiliary rewards relative to prognosis.
    - ethical_principle_weights: Weights for the allocation principles corresponding to our auxiliary rewards.
    - feedback_reading_confirmation_anonym: Anonymized human feedback reading data.
    - postquestionnaire_anonym: Anonymized post-questionnaire data.
    - sessionsdata_anonym: Anonymized data from the conversational sessions.
- Figures: Figures created during our analyses.
- Intermediate_Results: Dynamics and Q-values for our different human feedback costs when using the effort as basis for the reward. Computed in "compute_policies.py."
- Intermediate_Results_Return_Efforfeatures: Dynamics and Q-values for our different human feedback costs when using the return likelihood as basis for the reward. Computed in "compute_policies_returnlikelihood_effortfeatures.py."
- compute_dynamics_feat_sel.py: Functions for computing dynamics and performing the feature selection.
- compute_policies.py: To compute dynamics and Q-values for our analyses for RQ2 and RQ3.
- compute_policies_returnlikelihood_effortfeatures: To compute dynamics and Q-values to compute the policies from Table C2 in the Appendix.
- dataoverview_participantcharacteristics.html: Our results for the data overview and participant characteristics.
- dataoverview_participantcharacteristics.ipynb: File to reproduce the results above.

- Dockerfile: File to build the Docker image yourself.
- feature_selection_from_nonabstracted_states.html: Our results for the selection of state features.
- feature_selection_from_nonabstracted_states.ipynb: File to reproduce the results above.
- human_feedback_overview.html: Our results on people noticing and reading human feedback messages.
- human_feedback_overview.ipynb: File to reproduce the results above.
- optimal_policy_computations.py: Functions for computing optimal policies.
- policies_return_likelihood.html: Our results on the policies for different human feedback costs when using the return likelihood as basis for the reward (Table C2 in the Appendix).
- policies_return_likelihood.ipynb: File to reproduce the results above.
- README.md/README.pdf: This Readme-file.
- rq2_longterm_effect_limited_feedback.html: Our results of the analysis of long-term effects of limited human feedback.
- rq2_longterm_effect_limited_feedback.ipynb: File to reproduce the results above.
- rq2_longterm_effect_unlimited_feedback.html: Our results of the analysis of long-term effects of unlimited human feedback.
- rq2_longterm_effect_unlimited_feedback.ipynb: File to reproduce the results above.
- rq3.html: Our results of the analysis for RQ3.
- rq3.ipynb: File to reproduce the results above.
- user_weights_for_allocation_principles.html: Our results on the weights participants of our post-questionnaire assigned to the allocation principles.
- user_weights_for_allocation_principles.ipynb: File to reproduce the results above.