

Human-Robot Interactions

Assignment 1: Verbal Interaction

Lauren Kersten (s3950905) Yara Bikowski (s3989585)

March 4, 2024

1 Model

To implement human-like verbal interactions with our social robot, we decided to work with a pre-trained Large Language Model (LLM) as this would give more natural and novel language generation as opposed to scripted conversations or implementing a dialogue manager such as a POMDP.

At first, we tried the Qwen Chat model (i.e. `Qwen/Qwen1.5-0.5B-Chat`) [Bai et al., 2023]. However, we found its dialogue to not be very “human-like” based on its responses clarifying that it’s an AI model that had no preferences or consciousness. Thus we opted to use a different model, which is the `GODEL` model by Microsoft [Peng et al., 2022] (more specifically, `GODEL-v1.1-large-seq2seq`). We use this model as our final model as we found the conversation to be quite human-like in the way it forms its sentences and responses. Furthermore, it continues the conversation if the conversation dies down or the user says something ambiguous like “hmm”. An example dialogue can be found in Appendix A.

This is a Transformer-based encoder-decoder model trained on 551 million multi-dialogue Reddit discussions and 5 million instruction and knowledge-grounded dialogs. It makes use of an instruction (prompt) and a knowledge base. The instruction that we have chosen is the following: “Given a dialog context, you need to respond happily, friendly, empathically and easily understandable.” Through experimentation, we have found that this prompt encourages the model to talk about stuff that it and the user enjoy. The knowledge base that this model uses can further be used to make the robot aware of information like appointments that an elderly person has by mentioning the knowledge during conversation. We haven’t used this as it focuses the conversation too much on the knowledge which we did not want for our current implementation.

To add some novelty to the robot’s greetings and sign-offs, we created lists of different ‘Hellos’ and ‘Goodbyes’ of which the robot will randomly select one to say to the user so that it won’t repeatedly say the same greeting. The conversation does not have a finite length of interactions but when the user says something like “Goodbye”; the robot will sign off and terminate the conversation.

2 Evaluation Study

For our experiment, we would have liked to have three fellow students converse with our robot. These students would be instructed to talk with the robot about their hobbies and activities as if they were having a conversation with one of their friends. This instruction is only given to provide a starting point for the conversation and thus would not be enforced during the experiment. The experiment would have a maximum duration of 10 minutes for each participant. However, participants can choose to end the experiment earlier if they don’t want to talk with the robot anymore.

To evaluate the performance of our robot, we ask the participants to fill out a post-experiment survey where they can evaluate different aspects of the interaction with the robot. This survey consists of both open and closed questions, where the closed questions all made use of a 5-point Likert scale. Based on the quantitative and qualitative feedback from the survey, we would then evaluate the robot’s interaction

as well as make any possible improvements.

Survey questions

1. (open question) Did you end the conversation with the robot early? If yes, give a short explanation for why.
2. (closed question) I found the robot friendly.
3. (closed question) The conversation seemed human-like.
4. (closed question) I found the conversation enjoyable.
5. (closed question) Aside from potential technical limitations, the verbal interaction was implemented well.
6. (closed question) It was easy to have a conversation with the robot.
7. (open question) What is the robot’s biggest weakness?
8. (open question) Please provide any points for improvement you can think of.

References

- [Bai et al., 2023] Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., Hui, B., Ji, L., Li, M., Lin, J., Lin, R., Liu, D., Liu, G., Lu, C., Lu, K., Ma, J., Men, R., Ren, X., Ren, X., Tan, C., Tan, S., Tu, J., Wang, P., Wang, S., Wang, W., Wu, S., Xu, B., Xu, J., Yang, A., Yang, H., Yang, J., Yang, S., Yao, Y., Yu, B., Yuan, H., Yuan, Z., Zhang, J., Zhang, X., Zhang, Y., Zhang, Z., Zhou, C., Zhou, J., Zhou, X., and Zhu, T. (2023). Qwen technical report. *arXiv preprint arXiv:2309.16609*.
- [Peng et al., 2022] Peng, B., Galley, M., He, P., Brockett, C., Liden, L., Nouri, E., Yu, Z., Dolan, B., and Gao, J. (2022). Godel: Large-scale pre-training for goal-directed dialog.

A Example Dialog

This is an example dialog that shows how the robot can carry on the conversation even when it does not hear the user input entirely correctly. Below, you can see that when the ASR hears “feel like” instead of the correct response it begins to ask more questions about the user such as what their hobbies are. Although sometimes the responses are a bit random such as the comment about the nurses, it is its way of carrying on the conversation.

Dialog:

```
Robot: 'Good Afternoon!',
User: 'hello',
Robot: 'hello! how are you? ',
User: 'good how are you',
Robot: 'how are you? hello! you?',
User: "I'm okay how are you",
Robot: 'are you still alive? I'm not even alive!',
User: "what's your favorite color",
Robot: 'What is your favorite color? I love red.',
User: 'also like green',
Robot: 'What is your favorite food? Green is my favorite.',
User: 'feel like',
Robot: 'what are your hobbies? You have been a nurse'
```