A MultiModal Social Robot Toward Personalized Emotion Interaction

Baijun Xie and Chung Hyuk Park
Department of Biomedical Engineering
School of Engineering and Applied Science
The George Washington University, U.S.A.
bdxie@gwu.edu, chpark@gwu.edu

Abstract

Human emotions are expressed through multiple modalities, including verbal and non-verbal information. Moreover, the affective states of human users can be the indicator for the level of engagement and successful interaction, suitable for the robot to use as a rewarding factor to optimize robotic behaviors through interaction. This study demonstrates a multimodal human-robot interaction (HRI) framework with reinforcement learning to enhance the robotic interaction policy and personalize emotional interaction for a human user. The goal is to apply this framework in social scenarios that can let the robots generate a more natural and engaging HRI framework.

Introduction

This paper presents an ongoing study on multimodal human-robot interaction (HRI) with a reinforcement learning (RL) dialogue agent. To develop an effective HRI system for social robots that can naturally interact with human users, the robots need to accurately identify the user's affective states and respond to the users with personalized behaviors accordingly. Moreover, robots can also improve the user's engagement in the long-term interaction (Leite, Martinho, and Paiva 2013) through personalized interaction policy. However, in reality, emotion recognition could be challenged in HRI since humans convey information and feelings via different sources of social cues such as facial expression, body language, and speech. Thus, the studies of multimodal emotion recognition have attracted considerable interest in recent years.

These fusion strategies in the previous years can be summarized into three main categories: feature-level fusion, decision-level fusion, and model-level fusion (Wu, Lin, and Wei 2014). Schuller et al. (2012) demonstrated a baseline model in the Audio/Video Emotion Challenge (AVEC) 2021, which concatenated the audio and visual features using feature-level fusion and then used support vector regression to predict the continuous affective values. On the other hand, the decision-level fusion can process different types of inputs with diverse classifiers, and the final estimation can fuse the outputs from all classifiers. For example, the posterior probabilities from the predictions of the audio and video classifiers can be combined to get the final estimations (Schuller et al. 2011). Deep learning models, such as Long-Short Term Memory Recurrent Neural Network (LSTM-RNN), have been used to achieve a model-level fusion (Chen and Jin 2016). More recently, researchers have leveraged the transformer to fuse different modalities of inputs (Tsai et al. 2019; Xie, Sidulova, and Park 2021), where the transformer is proposed by Vaswani et al. (2017) for solving Sequence-to-Sequence (Seq2Seq) problem based on attention mechanism only without any recurrent structure as RNN.

Once the robots have the ability to recognize the user's affective states, the HRI system can utilize this information as inputs and determine the behaviors of the robots. The utilization of emotional models in HRI can create more natural and engaging HRI experiences, as evidenced by Ficocelli et al. (Ficocelli, Terao, and Nejat 2015). The developed emotional model can also be used for the empathetic appraisal for social robots that can interact with children in the long-term study (Leite et al. 2014). One recent study also presents a social robot with emotional interaction that can elicit particular emotions using non-verbal emotional behaviors (Shao et al. 2020). Thus, we believe that the affective states of the human user can be an effective indicator for the HRI system to generate more engaging and natural behaviors for the interaction.

Moreover, in order to develop a more natural HRI framework, the robots need to learn human preferences and skills. Recent studies utilize RL to let the robots learn social skills through the interactions (Kim et al. 2017; Qureshi et al. 2018). Furthermore, Ritschel (2018) proposed an RL framework that can adapt robot behavior using social signals. Robots' capacity to learn multimodal cues adaptively and associate them with the context is recognized as being the vital factor in HRI. Cui et al. (2020) proposed a novel data-driven framework, EMPATHIC, which aims to improve the policy of the agent in learning tasks by using implicit human facial reactions feedback. However, there is still a need for the study on the RL agent for multimodal HRI. Thus, in this study, we present a multimodal HRI system with an RL framework that can accept multiple modalities of inputs to shape the reward signal and adapt the robot behaviors to generate more positive feedback for personalized interaction

with the user. For the rest of the paper, We first present the developed dialogue RL framework in our system cause the speech will be the primary social cue for our robot platform. Then, we depict the integral diagram of our multimodal HRI system. Finally, we propose the evaluation plan for our future study.

Related Studies

Recent studies in multimodal HRI are providing enhanced abilities for emotional interaction (Liu et al. 2016; Hong et al. 2020; Li et al. 2019), where the emotional abilities, including understanding users' affective states and expressing emotional behaviors accordingly. In the study proposed by Liu et al. (2016), multiple modalities, including vocal, facial, and gesture features, were combined to detect the emotions by using a Bayesian classifier. The Nao humanoid robot was used as the platform to convey emotional behaviors by mimicking the user's behaviors. Hong et al. (2020) proposed a multimodal emotional HRI architecture that can interpret human emotional states via multimodal social cues and promote natural bidirectional communications with users. Gestures and vocal features were fused to detect the user's affect and the robot utilized the user's affect to express the corresponding emotional behaviors. Another study by Li et al. (2019) also used emotions for a spoken dialogue system in multimodal HRI which aimed to conduct the natural conversation. The study proposed an emotion recognition model which combines valence from prosody and sentiment analysis for the robot to express reactive emotions adequately. However, previous studies failed to consider using a dedicated dialogue system for the HRI system (Liu et al. 2016; Hong et al. 2020) or did not use different multimodal reactive emotion expressions for the robot (Li et al. 2019).

Methodology

This section will first formulate the natural language processing (NLP) problem we will solve with the RL agent based on physiological rewards. Then, we will introduce our overall robot system that can personalize the behaviors of our robot for more natural HRI scenarios.

Problem Formulation

In the Seq2Seq problem of NLP, the goal is to train a language model (LM) than can create a mapping between the input sequence $(x_0, \ldots, x_n) \in \mathcal{X}$ and output sequence $(y_0, \ldots, y_n) \in \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output space, respectively. Given a vocabulary $\sum$, where both $\mathcal{X}$ and $\mathcal{Y} \in \sum$, and a LM can generate sequences of tokens $(y_0, \ldots, y_n) \in \mathcal{Y}$ with a probability distribution using the chain rule (Bengio et al. 2003):

$$p(y_0, \ldots, y_n) = \prod_{0 \le i < n} p(y_i | x_0, \ldots, x_{n-1}). \quad (1)$$

Given the current sequences of tokens, LM can generate the probability distribution of the next token. We use the probability distribution assigned by LM as the initialization of the policy, $\pi_\theta = p$ with learnable parameters θ, and then train π_θ via RL. A reward function $r : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ can be defined and use RL to optimize the expected reward:

$$\mathbb{E} = \mathbb{E}_{x,y}[r(x, y)]. \quad (2)$$

We formulate the reward r as a distance-based function based on Russel's circumplex psychological model of affect (Russell 1980). The positive emotions, such as *happy*, and *excited* have positive rewards given by the values of arousal (A) and valence (V) of these emotions. On the contrary, the negative feelings will be assigned negative reward values. Thus, the reward can be expressed as follows:

$$r(x, y) = \pm\sqrt{A^2 + V^2}. \quad (3)$$

The objective of our RL task is to generate more positive responses. To achieve this, we need first to train a reward model to estimate the reward based on the generated embedding vectors from the LM. Following the setting of the previous study (Xie and Park 2021), a linear classifier is added on the top of the LM to estimate the emotion and predict the level of arousal and valence values. The reward model is optimized using this loss:

$$\mathcal{L}_r = \mathbb{E}_{x,y}[r(x, y)]. \quad (4)$$

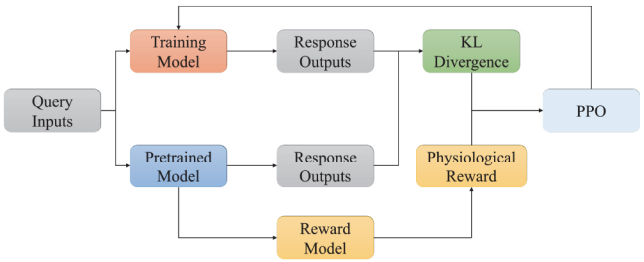

Figure 1: The workflow of the proximal policy optimization (PPO) with psychological rewards.

Following Ziegler et al. (2019), when fine-tuning the policy π_θ to optimize the reward via proximal policy optimization (PPO) algorithm (Schulman et al. 2017), the Kullback-Leibler (KL) divergence penalty is added to prevent π_θ from changing too fast from p. Thus, a modified reward function is given by:

$$R(x, y) = r(x, y) - \beta \log \frac{\pi_\theta(y|x)}{p(y|x)}, \quad (5)$$

where β can be chosen either fixed or adaptive using KL values (Schulman et al. 2017). As can be seen in Figure 1, the main objective for policy iteration will consider both psychological rewards and the KL values. The overall training process is given in Algorithm 1. The RL agent will optimize the LM for each sentence with annotated emotions, and by leveraging the reward function 5, the LM model will generate more positive responses but doesn't deviate too much compared with the original model.

Algorithm 1: PPO with psychological reward

Algorithm parameters: either fixed or dynamic β;
Initialize $\pi = p$;
Initialize r to p (reward model added on top of LM);
foreach *episode* **do**
  Run language model policy π to generate sentence embeddings vectors;
  Optimize the reward model by the embedding vectors using loss (4);
  Update the language model parameters θ via PPO with reward function (5);
  $\theta \leftarrow \theta'$;
end
return π^*;

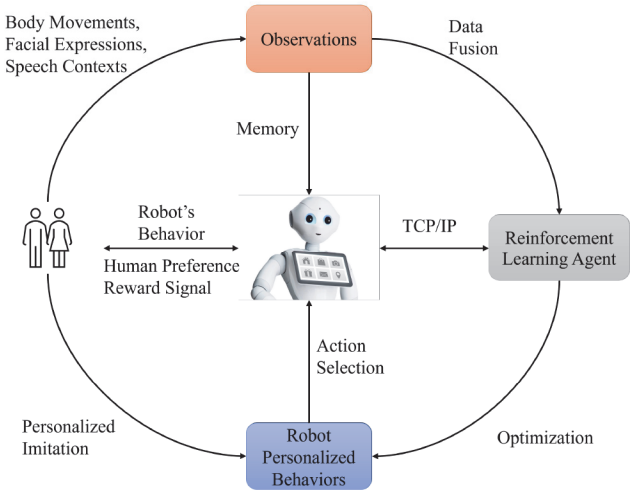

Figure 2: The proposed multimodal HRI framework with RL agent. Our robotic system (Pepper) will observe and learn human behaviors and multimodal responses, while updating behavioral policy focused on personalizing for each user.

Pre-trained Language Model

Generative Pre-training Transformer (GPT) (Radford et al. 2018) is considered to be employed as the pre-trained LM in this study. GPT is transformer-based LM, where the transformer is designed purely based on the attention mechanism proposed by Vaswani et al. (2017). GPT consists of multiple layers of transformer with self-attention operations, and it is also pre-trained on the large corpus.

To enable the LM the ability to recognize emotions from the dialogue, the GPT is also fine-tuned on a dialogue dataset, MELD dataset (Poria et al. 2018), which is targeting for the task of emotion recognition during the conversation. Each sentence from each dialogue sample of the dataset is annotated with an emotion category. The emotion labels available in the MELD dataset are *Anger*, *Disgust*, *Sadness*, *Joy*, *Neutral*, *Surprise*, and *Fear*. The expert annotations of this dataset can provide the model the ability to assess the emotion of each user's response. The reward model, r in the RL optimization process, is added as an extra layer on top of the GPT model, which is also mentioned in Algorithm 1.

Human-Robot Interaction Design

A multimodal HRI system will be designed for evaluating the RL framework. As can be seen in Figure 2, several modalities of observations, such as body gestures, facial expression, and speech, will be considered as the social cues and input to the system. The extracted features for the body gestures will be the human skeleton joint poses, and facial landmarks for the facial expression, and sentence embeddings for the conversational speech. The robotic platform will be the Pepper robot manufactured by the SoftBank robotics group (SoftBank 2021). Several modalities of inputs are then to be fused as part of the reward signals to the RL agent. The overall reward signal will also combine the human preferences, such as the self-evaluations from the user as the intrinsic rewards. We will adopt a self-assessment scale for the measurement of user's emotions, e.g., the self-assessment manikin (Bradley and Lang 1994), which is a non-verbal assessment tool that can measure affective reactions to different stimuli. The personalized robot behaviors are learned by creating the mapping between the interaction context and the affective states. The interaction context includes the textual, vocal, and gesture information from the users. The personalized robot behaviors will also be optimized through the RL reward signals.

Evaluation Plan

We plan to conduct a user study to evaluate the performance of our HRI system in generating personalized behaviors and eliciting positive emotions from the users. A two-stage study will be conducted. In the first stage, we will give the model the general knowledge of recognizing human emotions via multimodal inputs and generating feedback accordingly. To achieve this, we will use a large-scale dataset to train the model. Moreover, we will test the ability of the LM to generate positive responses after optimization by the reward signals. In the second stage, we will hold the user study to evaluate our HRI system. The agent will be personalized by utilizing the user's multimodal information and preferences. Our hypothesis is that multimodal social cues can give the robot a better ability to understand human emotional states and enhance interactive behaviors. To evaluate the user study, we propose to use Negative Attitudes towards Robots Scale (NARS) (Nomura et al. 2006) and Engagement (Sidner et al. 2004) questionnaires. This study hypothesizes that the proposed measures between the test group and the control group should be significantly different. A two-sided independent samples t-test would be done to compare the mean values of every extracted feature between two groups.

The experiment will be held using the Pepper humanoid robot. The user can prompt different basic topics to express their emotions during the interaction, and our empathetic dialogue agent in the system will respond accordingly. The built-in dialogue module from the Pepper robot will be used as the baseline for the control group, which only includes

simple pre-defined chit-chat and rule-based contents. After the experiment, the NARS and Engagement questionnaires will be used to score both models.

Conclusion

In this study, we present a multimodal HRI framework with an RL agent. The NLP problem optimized by RL is formulated and solved by the PPO algorithm. The physiological measures will be investigated to be used as the reward signals for optimizing the RL agent. GPT is used as the pre-trained language model for the dialogue system. The proposed multimodal HRI framework will be evaluated with a two-stage user study. Different inputs modalities will be fused to recognize the user's affective states, which will also be treated as intrinsic reward signals for the RL agent. By fusing the multimodal behaviors/responses and preferences as rewards from the users, the robotic behaviors can be personalized through learning human skills and preferred feedback. We are optimistic that this work will extend the current research on social robots to provide more natural and personalized interaction capabilities, especially using multimodal HRI interaction policies and optimization through RL. For the future study, we propose using this framework to analyze and detect emotional cues from multimodal interaction data and provide personalized intervention for adolescents with Autism Spectrum Disorder (ASD).

Acknowledgment

This research is supported by the National Science Foundation (NSF) under the grant #1846658.

References

Bengio, Y.; Ducharme, R.; Vincent, P.; and Janvin, C. 2003. A neural probabilistic language model. *The journal of machine learning research* 3: 1137–1155.

Bradley, M. M.; and Lang, P. J. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25(1): 49–59.

Chen, S.; and Jin, Q. 2016. Multi-modal conditional attention fusion for dimensional emotion prediction. In *Proceedings of the 24th ACM international conference on Multimedia*, 571–575.

Cui, Y.; Zhang, Q.; Allievi, A.; Stone, P.; Niekum, S.; and Knox, W. B. 2020. The empathic framework for task learning from implicit human feedback. *arXiv preprint arXiv:2009.13649* .

Ficocelli, M.; Terao, J.; and Nejat, G. 2015. Promoting interactions between humans and robots using robotic emotional behavior. *IEEE transactions on cybernetics* 46(12): 2911–2923.

Hong, A.; Lunscher, N.; Hu, T.; Tsuboi, Y.; Zhang, X.; dos Reis Alves, S. F.; Nejat, G.; and Benhabib, B. 2020. A Multimodal Emotional Human-Robot Interaction Architecture for Social Robots Engaged in Bidirectional Communication\vspace* 7pt. *IEEE transactions on cybernetics* .

Kim, S. K.; Kirchner, E. A.; Stefes, A.; and Kirchner, F. 2017. Intrinsic interactive reinforcement learning–Using error-related potentials for real world human-robot interaction. *Scientific reports* 7(1): 1–16.

Leite, I.; Castellano, G.; Pereira, A.; Martinho, C.; and Paiva, A. 2014. Empathic robots for long-term interaction. *International Journal of Social Robotics* 6(3): 329–341.

Leite, I.; Martinho, C.; and Paiva, A. 2013. Social robots for long-term interaction: a survey. *International Journal of Social Robotics* 5(2): 291–308.

Li, Y.; Ishi, C. T.; Inoue, K.; Nakamura, S.; and Kawahara, T. 2019. Expressing reactive emotion based on multimodal emotion recognition for natural conversation in human–robot interaction. *Advanced Robotics* 33(20): 1030–1041.

Liu, Z.-T.; Pan, F.-F.; Wu, M.; Cao, W.-H.; Chen, L.-F.; Xu, J.-P.; Zhang, R.; and Zhou, M.-T. 2016. A multimodal emotional communication based humans-robots interaction system. In *2016 35th Chinese Control Conference (CCC)*, 6363–6368. IEEE.

Nomura, T.; Suzuki, T.; Kanda, T.; and Kato, K. 2006. Measurement of negative attitudes toward robots. *Interaction Studies* 7(3): 437–454.

Poria, S.; Hazarika, D.; Majumder, N.; Naik, G.; Cambria, E.; and Mihalcea, R. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508* .

Qureshi, A. H.; Nakamura, Y.; Yoshikawa, Y.; and Ishiguro, H. 2018. Intrinsically motivated reinforcement learning for human–robot interaction in the real-world. *Neural Networks* 107: 23–33.

Radford, A.; Narasimhan, K.; Salimans, T.; and Sutskever, I. 2018. Improving language understanding by generative pre-training .

Ritschel, H. 2018. Socially-aware reinforcement learning for personalized human-robot interaction .

Russell, J. A. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39(6): 1161.

Schuller, B.; Valstar, M.; Eyben, F.; McKeown, G.; Cowie, R.; and Pantic, M. 2011. Avec 2011–the first international audio/visual emotion challenge. In *International Conference on Affective Computing and Intelligent Interaction*, 415–424. Springer.

Schuller, B.; Valster, M.; Eyben, F.; Cowie, R.; and Pantic, M. 2012. Avec 2012: the continuous audio/visual emotion challenge. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, 449–456.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* .

Shao, M.; Snyder, M.; Nejat, G.; and Benhabib, B. 2020. User affect elicitation with a socially emotional robot. *Robotics* 9(2): 44.

Sidner, C. L.; Kidd, C. D.; Lee, C.; and Lesh, N. 2004. Where to look: a study of human-robot engagement. In *Proceedings of the 9th international conference on Intelligent user interfaces*, 78–84.

SoftBank. 2021. Pepper the humanoid and programmable robot: SoftBank Robotics, Accessed 13 August 2021. URL https://www.softbankrobotics.com/emea/en/pepper.

Tsai, Y.-H. H.; Bai, S.; Liang, P. P.; Kolter, J. Z.; Morency, L.-P.; and Salakhutdinov, R. 2019. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the conference. Association for Computational Linguistics. Meeting*, volume 2019, 6558. NIH Public Access.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008.

Wu, C.-H.; Lin, J.-C.; and Wei, W.-L. 2014. Survey on audiovisual emotion recognition: databases, features, and data fusion strategies. *APSIPA transactions on signal and information processing* 3.

Xie, B.; and Park, C. H. 2021. Empathetic Robot With Transformer-Based Dialogue Agent. In *2021 18th International Conference on Ubiquitous Robots (UR)*, 290–295. IEEE.

Xie, B.; Sidulova, M.; and Park, C. H. 2021. Robust Multimodal Emotion Recognition from Conversation with Transformer-Based Crossmodality Fusion. *Sensors* 21(14): 4913.

Ziegler, D. M.; Stiennon, N.; Wu, J.; Brown, T. B.; Radford, A.; Amodei, D.; Christiano, P.; and Irving, G. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593* .