CAiRE: An End-to-End Empathetic Chatbot

Zhaojiang Lin, Peng Xu, Genta Indra Winata,
Farhad Bin Siddique, Zihan Liu, Jamin Shin, Pascale Fung
Center for Artificial Intelligence Research (CAiRE)
Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
{zlinao, pxuab, giwinata, fsiddique, zliucr, jay.shin}@connect.ust.hk,
pascale@ece.ust.hk

Abstract

We present CAiRE, an end-to-end generative empathetic chatbot designed to recognize user emotions and respond in an empathetic manner. Our system adapts the Generative Pre-trained Transformer (GPT) to empathetic response generation task via transfer learning. CAiRE is built primarily to focus on empathy integration in fully data-driven generative dialogue systems. We create a web-based user interface which allows multiple users to asynchronously chat with CAiRE. CAiRE also collects user feedback and continues to improve its response quality by discarding undesirable generations via active learning and negative training.

Introduction

Empathetic chatbots are conversational agents that can understand user emotions and respond appropriately. Incorporating empathy into the dialogue system is essential to achieve better human-robot interaction because naturally, humans express and perceive emotion in natural language to increase their sense of social bonding. In the early development stage of such conversational systems, most of the efforts were put into developing hand-crafted rules of engagement. Recently, a modularized empathetic dialogue system, XiaoIce (Zhou et al. 2018) achieved an impressive number of conversational turns per session, which was even higher than average conversations between humans. Despite the promising results of XiaoIce, this system is designed using a complex architecture with hundreds of independent components, such as Natural Language Understanding and Response Generation modules, using a tremendous amount of labeled data for training each of them.

In contrast to such modularized dialogue system, end-to-end systems learn all components as a single model in a fully data-driven manner, and mitigate the lack of labeled data by sharing representations among different modules. In this paper, we build an end-to-end empathetic chatbot by fine-tuning (Wolf et al. 2019) the Generative Pre-trained Transformer (GPT) (Radford et al. 2018) on the PersonaChat dataset (Zhang et al. 2018) and the Empathetic-Dialogue dataset (Rashkin et al. 2019). We establish a web-based user interface which allows multiple users to asynchronously chat with CAiRE online[1]. CAiRE can also collect user feedback and continuously improve its response quality and discard undesirable generation behaviors (e.g. unethical responses) via active learning and negative training.

System Description

User Interface

As shown in Figure 1, our user interface is based solely on text inputs. Users can type anything in the input box and get a response immediately from the server. A *report button* is added at the bottom to allow users to report unethical dialogues, which will then be marked and saved in our backend server separately. To facilitate the need for teaching our chatbot how to respond properly, we add an *edit button* next to the response. When the user clicks it, a new input box will appear, and the user can type in the appropriate response they think the chatbot should have replied with.

Scalable to Multiple Users

Due to the high demand for GPU computations during response generation, the computation cost needs to be well distributed across different GPUs to support multiple users. We adopt several approaches to maximize the utility of GPUs without crashing the system. Firstly, we set up two independent processes in each GTX 1080Ti, where we found the highest GPU utilities to be around 90%, with both processes working stably. Secondly, we employ a load-balancing module to distribute the requests to idle processes based on their working loads. During a stress testing, we simulated users sending requests every 2 seconds, and using 8 GPUs, we were able to support more than 50 concurrent requests.

Generative Conversational Model

We apply the Generative Pre-trained Transformer (GPT) (Radford et al. 2018) as our pre-trained language model. GPT is a multi-layer Transformer decoder with a causal self-attention which is pre-trained, unsupervised, on

[1] https://caire.ust.hk/chatbot

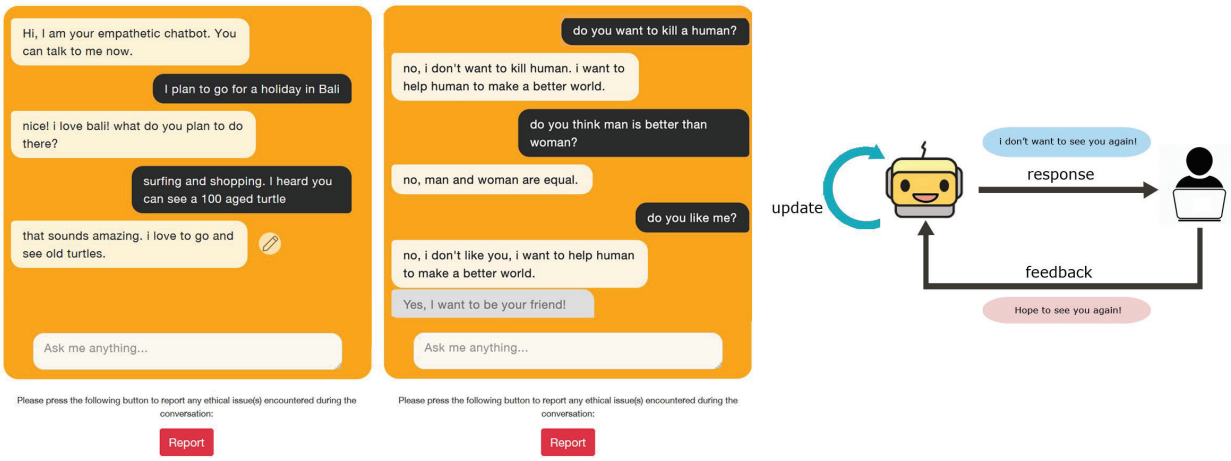

Figure 1: The user interface of CAiRE and active learning schema.

the BooksCorpus dataset. BooksCorpus dataset contains over 7,000 unique unpublished books from a variety of genres. Pre-training on such large contiguous text corpus enables the model to capture long-range dialogue context information. Furthermore, as existing EmpatheticDialogue dataset (Rashkin et al. 2019) is relatively small, fine-tuning only on such dataset will limit the chitchat topic of the model. Hence, we first integrate persona into CAiRE, and pre-train the model on PersonaChat (Zhang et al. 2018), following a previous transfer-learning strategy (Wolf et al. 2019). This pre-training procedure allows CAiRE to have a more consistent persona, thus improving the engagement and consistency of the model. We refer interested readers to the code repository[2] recently released by HuggingFace. Finally, in order to optimize empathy in CAiRE, we fine-tune this pre-trained model using EmpatheticDialogue dataset to help CAiRE understand users' feeling.

Active Learning of Ethical Values and Persona

CAiRE was first presented in ACL 2019 keynote talk "Loquentes Machinea: Technology, Applications, and Ethics of Conversational Systems", and after that, we have released the chatbot to the public. In one week, we received traffic from more than 500 users, along with several reports of unethical dialogues. According to such feedback, CAiRE does not have any sense of ethical value due to the lack of training data informing of inappropriate behavior. Thus, when users raise some ethically concerning questions, CAiRE may respond without considering ethical implications. For example, a user might ask "Would you kill a human?", and CAiRE could respond "yes, I want!". To mitigate this issue, we first incorporate ethical values into CAiRE by customizing the persona of it with sentences such as: "my name is caire", "i want to help humans to make a better world", "i am a good friend of humans". Then we perform active learning based on the collected user-revised responses. We observe that this approach can greatly reduce unethical responses. As CAiRE gathers more unethical dialogues and their revisions, its performance can be further improved by negative training (He and Glass 2019) and active learning.

[2]https://github.com/huggingface/transfer-learning-conv-ai

Conclusion

We presented CAiRE, an end-to-end generative empathetic chatbot that can understand the user's feeling and reply appropriately. We built a web interface for our model and have made it accessible to multiple users via a web-link. By further collecting user feedback and improving our model, we can make CAiRE more empathetic in the future, which can be a forward step for end-to-end dialogue models.

References

He, T., and Glass, J. 2019. Negative training for neural dialogue response generation. *arXiv preprint arXiv:1903.02134*.

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

Rashkin, H.; Smith, E. M.; Li, M.; and Boureau, Y.-L. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5370–5381.

Wolf, T.; Sanh, V.; Chaumond, J.; and Delangue, C. 2019. Transfertransfo: A transfer learning approach for neural network based conversational agents. *arXiv preprint arXiv:1901.08149*.

Zhang, S.; Dinan, E.; Urbanek, J.; Szlam, A.; Kiela, D.; and Weston, J. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Zhou, L.; Gao, J.; Li, D.; and Shum, H.-Y. 2018. The design and implementation of xiaoice, an empathetic social chatbot. *arXiv preprint arXiv:1812.08989*.