The Empathic Dialogue Generation Model Based on Emotion Cause Perception

Yun Su[1,2,*,†], Bozhen Fan[1,†], Haoran Bian[1,†], Runhe Huang[2,†] and Yunhao Zhu[1,†]

[1]Northwest Normal University of China , Lanzhou ,730070, China
[2]Hosei University,Tokyo, 1848584, Japan

Abstract

Current methods for generating empathy dialogues often overlook the emotional triggers that lead to changes in emotions. To address this issue, we present a novel framework that enhances empathetic response generation by identifying emotional causes within conversations. Our framework consists of two modules: one that comprehends emotions originating from both content and context, and another that features an emotional attention mechanism for empathy expression. Experimental results demonstrate that our proposed model is capable of perceiving emotional causes and can improve the quality of empathy expression.

Keywords

emotional conversation generation, emotion cause detection, empathetic response generation

1. Introduction

The perception and expression of emotion is very important to the generation of dialogue. Emotional causes are events that trigger changes in the speaker's emotions . Failure to analyze emotional causes could lead to poor emotional perception [1]. To address this issue, we propose a framework that improves the generation of empathetic responses by endowing the empathetic dialogue model with the ability to reason about human emotions in conversations. Our framework comprises two components: an emotion reasoner and a response generator. The experimental results show that our proposed model outperforms other compared methods by considering emotional causes in generating more empathetic responses.

2. Approach

Our model architecture is illustrated in Figure 1, and it consists of two main modules: the emotion reasoner and the response generator. The first module, the emotion reasoner, is responsible for predicting both the context of the emotion cause and the corresponding emotion tag. The second module, the response generator, integrates the information provided by the emotion reasoner to generate an appropriate response.

For the emotion reasoning, two encoders, that is, semantic and emotional encoders are employed to understand the conversation context from both a content and emotional perspective and locate the words related to emotional causes.

The semantic encoder is used to process the historical dialogue input, which is denoted as $U = [CLS, x_1, x_2, ..., x_n]$, $[CLS]$ is a semantic classifier token. For each word x_i in the input, the semantic encoder assigns a word embedding vector $e^W_{x_i}$, a position embedding vector $e^P_{x_i}$, and a conversation state embedding vector $e^D_{x_i}$, which capture the semantic information, location in the context, and interlocutor information of each word, respectively. The obtained final context representations are denoted as $C_U = [\tilde{CLS}, \tilde{x}_1, \tilde{x}_, ..., \tilde{x}_n]$.

Similarly, the emotional encoder is used to process emotional words in the semantic context U, which is denoted as $E = [LAB, w_1, w_2, ..., w_e]$, Similar to $[CLS]$, $[LAB]$ is an emotion classifier token. For each emotional word w_i in the input, the encoder assigns a word embedding vector $e^W_{w_i}$, a position embedding vector $e^P_{w_i}$, and an emotional state embedding vector $e^W_{w_i}$, which capture the emotional information associated with each word. Then the multi-resolution emotional context is represented as $C_E = [\tilde{LAB}, \tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_e]$.

To perceive the emotional information in dialogue context, a linear layer with softmax operation projects the concatenation of w_0 and x_0 into an emotion category distribution P_e over the coarsed-grained emotional label e to identify the emotion signal user expressed:

$$P_e(e|\varepsilon) = softmax(W_e[\tilde{w}_0; \tilde{x}_0]) \tag{1}$$

The emotion cause detection is a sequence labelling problem. Each word in the sequence is labelled with an emotion cause-oriented label $\in \{0, 1\}$, indicating whether the word is related to the emotion caused. Then compute the probability c_i of the i-th word related to the emotion cause with a linear layer coupled with a softmax function:

$$P_e(c_i|v_{xi}) = softmax(W_e v_{xi} + b_c) \tag{2}$$

Note that the $[CLS]$ token is always labeled with 1. The sequence of emotion cause-oriented labels will later be used to select the emotion cause-related words in the input sequence to attend to for the response generator.

Finally, the two encoders are combined to generate the final dialogue representation $[Ce; Cu]$. At the same time, based on the semantic context vector representation, $C = [c_0, c_1, ..., c_n]$ is obtained through the full connection layer, and each word in the conversation context is assigned an emotional reason tag, where $c_i \in \{0, 1\}$. The tag sequence of emotional cause is marked to see whether each word in the conversation is the emotional reason word that causes the user's emotional changes so that the model can better understand the user's emotion caused by emotional reasons.

The emotion expression process is based on the decoder of the transformer. The emotion attention mechanism is set after the cross-attention mechanism so that the dialogue generation model can better focus on the emotion caused by user vector input. Then, the decoder exports the target vector $Y = [y_1, y_2, ..., y_j]$ from the dialogue context.

At the same time, to improve the ability of emotion recognition and semantic perception of the model, we only use the generated confrontation network in the training process. The discriminator part is inspired by [2]. It comprises two parts: emotion discriminator and semantic discriminator (discriminator part), as shown in the grey frame on the right in figure 1.

AAAI 2023 Spring Symposium, Socially Responsible AI for Well-being, March 27–29, 2023, USA
*Corresponding author.
†These authors contributed equally.
suyun@nwnu.edu.cn (Y. Su); 2021212117@nwnu.edu.cn (B. Fan); 2020222034@nwnu.edu.cn (H. Bian); rhuang@hosei.ac.jp (R. Huang); 2020222034@nwnu.edu.cn (Y. Zhu)
0000-0002-5906-4477 (B. Fan)

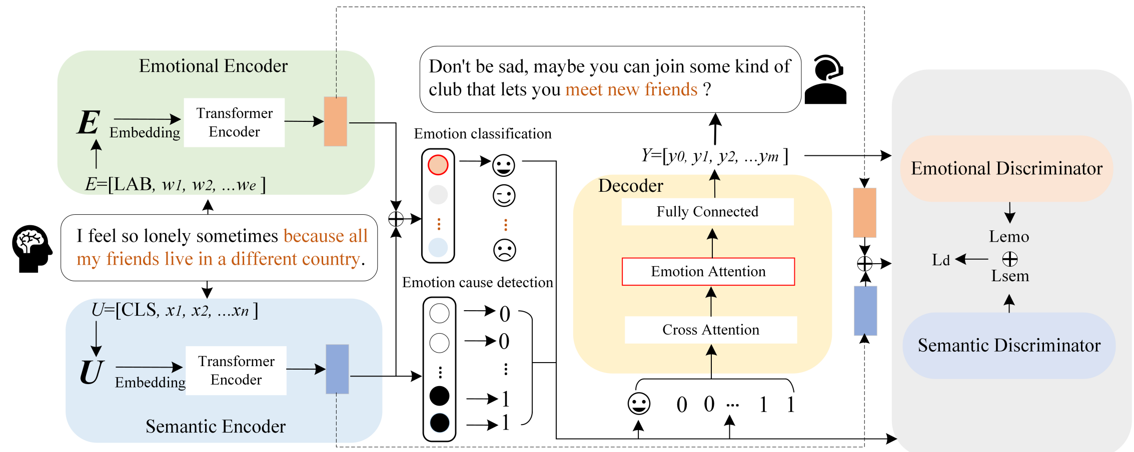

Figure 1: Architecture of the proposed model.

Table 1
Results on automatic evaluation and manual evaluation.

Models	Accuracy	Perplexity	Empathy	Revelence	Fluency
Moel	0.30	36.9	3.47	3.88	3.68
MIME	**0.32**	37.0	3.63	3.6	4.28
EMPDG	0.29	37.2	3.58	**3.91**	3.67
OUR	**0.32**	**33.4**	**3.77**	3.69	**3.73**

3. Experiments

Dataset To better capture the emotional content in user utterances, two different dataset are used: Empathtic Dialogues with emotional causes labels[3]. And the NRC Word-Emotion Association Lexicon (EmoLex) [4]. EmpatheticDialogues provides coarse-grained emotional labels for the dialogues, while EmoLex provides fine-grained emotional labels for individual words. The emotion cause is identified at the discourse level in the dialogues using an existing emotion cause detection model and label them accordingly in EmpatheticDialogues. This approach allows us to better understand the emotional context of the dialogues and provide more accurate emotional labels for the model training.

Baselines To assess our model effectiveness in capturing and generating empathetic responses with subtle emotional nuances, we compare our model's performance against several baselines, including the MoEL model [5], which is an extension of the Transformer model that combines response representations from different decoders optimized for different emotions; the MIME model [6] is another Transformer-based model that considers emotion clustering and emotional mimicry, and introduces sampling stochasticity during training; the EMPDG model [2] is a kind of empathic dialogue generation model based on generative adversarial network.

Evaluation Results As shown in table 1, our results have certain advantages in the accuracy of emotion recognition and the PPL of dialogue, which shows that our reasoning process on emotional causes helps the model to perceive emotion better, and at the same time, produces a more sympathetic expression. At the same accuracy rate of emotion recognition, our model has more advantages in the value of ppl, which shows that our model can better perceive the subtle emotional reasons and respond accordingly with the same recognition effect. These automatic evaluation results suggest that our approach is effective in generating empathetic responses with subtle emotional nuances and diverse language. At the same time, the results of the manual evaluation show that our empathy expression and fluency of sentences are also better.

4. Conclusion

The paper introduces a new framework that can enhance empathetic response generation by incorporating information about the causes of emotions. The evaluations demonstrate that the proposed models can generate more meaningful and empathetic responses compared to other existing approaches. By integrating emotional reasoning into conversation models, our framework has the potential to significantly improve the quality of human-computer interaction, particularly in scenarios where empathetic communication is essential.

5. Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61862058 and 8226070356), and in part by the China Scholarship Council(CSC).

References

[1] H. Herjanto, M. Amin, F. Okumus, and C. Cobanoglu, (2022). Airline service: Low-cost-carriers (LCCs) failure and passenger emotional experience. Tourism Review, 77(3), 945-963.

[2] Q. Li, H. Chen, Z. Ren, P. Ren, Z. Tu, Z. Chen, EmpDG:Multi-resolution Interactive Empathetic Dialogue Generation, Proc. 28th Int. Conf. Comput. Linguist. (2020). URL: https://doi.org/10.48550/arXiv.1911.08698

[3] J. Gao, Y. Liu, H. Deng, W. Wang, Y. Cao, J. Du, R. Xu, Improving Empathetic Response Generation by Recognizing Emotion Cause in Conversations, Find. Assoc. Comput. Linguist. Find. ACL EMNLP 2021. (2021) 807–819. URL: https://doi.org/10.18653/v1/2021.findings-emnlp.70.

[4] S. M. Mohammad, and P. D. Turney, (2013). Crowdsourcing a word–emotion association lexicon. Computational intelligence, 29(3), 436-465. URL: https://doi.org/10.1111/j.1467-8640.2012.00460.x

[5] Z. Lin, A. Madotto, J. Shin, P. Xu, and P. Fung, MOEL: Mixture of empathetic listeners, EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf. (2019) 121–132. URL: https://doi.org/10.48550/arXiv.1908.07687

[6] N. Majumder, P. Hong, S. Peng, J. Lu, and D. Ghosal, A. Gelbukh, R. Mihalcea, S. Poria, MIME: Mimicking emotions for empathetic response generation, in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), (2020), pp. 8968–8979. URL: https://doi.org/10.48550/arXiv.2010.01454