Emotional Conversation Generation based on Emotion-adaptive Transformer Network

A127

Dian Cheng and Shifeng Zhou
University of Science and Technology of China

1 Introduction

A good human-machine conversation system should not only focus on generating relevant and diverse responses, but also understand and process user's emotions. However, existing emotional conversation generation (ECG) approaches have not been able to incorporate emotional factors sufficiently into large-scale dialogue generation task while generating responses with good grammar and semantics. To tackle this limitation, this article proposes a new model that is fully applicable to ECG task on large-scale corpora. On the one hand, our model can finely encode semantic and emotional information in the input sentence. On the other hand, we explicitly model the expression process of emotions in decoding phase, constraining the model to generate high-quality responses that fully express emotions. Experiments conducted on two standard large-scale Chinese conversation datasets validate the superiority of our proposal over baseline models.

2 Methods

Our proposed ECG model consists of several components: semantic and emotional encoder modules, a decoder module, and emotional prediction modules, as shown in Figure 1. The model first passes the post to the semantic encoder and emotional encoder to compute the semantic context $\boldsymbol{M_s}$ and emotional context $\boldsymbol{M_e}$, and predicts the post emotion $\hat{\boldsymbol{e}}_p$ and the emotion that should be present in the response $\hat{\boldsymbol{e}}_r$ based on the emotional context. After generating a part of the response sentence $Y = (y_1, y_2, \dots, y_{n-1})$, the model incorporates the response emotion representation $\hat{\boldsymbol{e}}_r$ to enhance the emotion representation of Y and inputs it to the decoder along with $\boldsymbol{M_e}$ to generate predicted values for $(y_2, y_3, \dots, y_n)$. We will describe each component of the model in detail below.

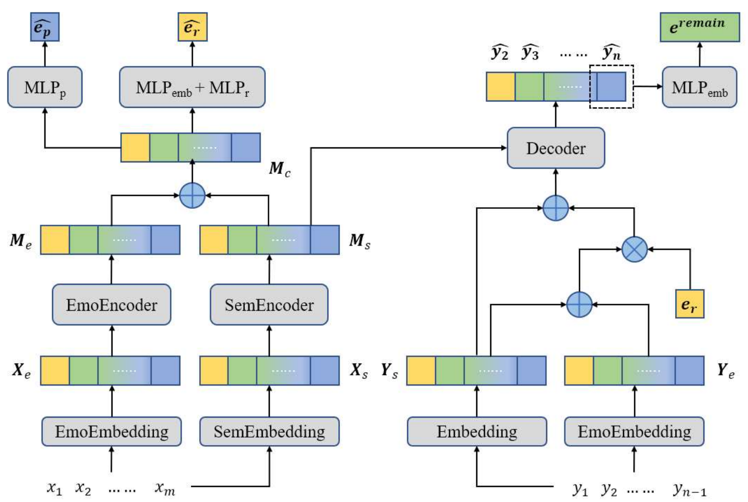

Figure 1. Architecture of our proposed model.

Table 1. Automatic evaluation results of each model.

Dataset	Rouge-1 (10-2)		Rouge-2 (10-2)		Distinct-1 (10-1)		Distinct-2 (10-1)	
	STC	NLPCC	STC	NLPCC	STC	NLPCC	STC	NLPCC
Seq2SeqA	0.101	0.064	0.050	0.016	0.032	0.036	0.133	0.067
ECM	0.191	0.069	0.063	0.019	0.035	0.029	0.153	0.065
EACM	0.173	0.047	0.142	0.010	0.030	0.031	0.124	0.059
BART	0.467	0.151	**0.404**	0.073	0.044	0.058	0.187	0.105
Ours	**0.503**	**0.155**	0.377	**0.074**	**0.049**	**0.062**	**0.212**	**0.117**

3 Discussion

First, we compare our model with ECM and EACM from the perspective of emotion factor incorporation method. ECM designs an external memory mechanism to control the selection process of emotion words. Our model embeds emotion information in independent emotion embedding module, without the need for a separate emotion vocabulary, making it more suitable for large-scale dialogue generation tasks. EACM simply passes the encoded emotion vector as part of the decoder input, without dynamically calculating the required emotion information at current decoding step. Our approach simulates the process of emotional expression, which is a more reasonable abstraction.

Then we compare from the perspective of encoding and decoding models. Existing ECG models based on LSTM and GRU require step-by-step encoding and decoding, which determines that it is relatively easy to extend LSTM/GRU to introduce emotional factors. In contrast, our model uses Transformer-based encoder and decoder, which has three advantages:

(1) It utilizes the excellent performance of the Transformer structure.

(2) In training procedure, the teacher forcing mechanism is used, which is more efficient without step-by-step decoding.

(3) Our proposed model can utilize the high-quality representations and weights learned by large-scale pre-trained language models to accelerate our training.

4 Conclusion

In this article, we propose a Transformer-based method for adaptive emotion expression in dialogue generation. This method models emotional representation at word level and can encode fine-grained emotions in posts. Additionally, we propose a dialogue generation method that models the process of emotional expression in decoding phase to generate responses with sufficiently rich emotion. Automatic evaluations and human evaluations on two real-world large-scale dialogue datasets demonstrate that our model produces text quality and emotional expression beyond baseline models.

Table 3. Case study on Seq2SeqA, ECM, BART and our model.

Post	Model / Emotion		Generated Response
Don't let your interest become its scourge.	Seq2SeqA		Hahahahahaha.
	ECM	happy	Hahahaha.
		like	I **like** the idea, but I can't do it.
		disgust	Isn't this our blood type?
		angry	It's disgusting, what is this stuff?
		sad	This is my life, even when I'm bored.
	BART		Don't let your interest become its scourge.
	Ours	happy	Hahaha, this is so funny, I **like** it.
		like	I think this picture is **beautiful**.
		disgust	What movie is this? !
		angry	**Oops**! What movie is this?
		sad	**Oh my god my heart is broken**.

Table 2. Manual evaluation result of selected models.

	STC	NLPCC
Seq2SeqA	0.40	0.37
ECM	0.51	0.52
BART	0.77	0.81
Ours	**0.79**	**0.86**