Emotional Response Generation in Multi-Turn Dialogue

Jinyao Yang1* and Chunhua Wu2
1 Cyber space security, Beijing University of Posts and Telecommunications, Beijing, Beijing, 100876, China
2 Cyber space security, Beijing University of Posts and Telecommunications, Beijing, Beijing, 100876, China
*Corresponding author’s e-mail: yangjinyao96@bupt.edu.cn

Abstract. The research on emotional response generation is one of the directions in dialog system research. However, most of the researches on emotional dialog generation focus on the expression of expected emotion in single round dialogue. In this paper, we propose a method to model the emotion in multi-turn dialogue. Our model bases on a proposed model: Emotional Chatting Machine, and solves the problem of ECM model that which emotion should be used to the response. We compared the effects of various models and finally chose LSTM to model emotions in multi-turn dialogue. Experiments show that the performance of the dialogue system can be effectively improved by modelling emotion in multi-turn dialogue.

1. Introduction
Dialogue system has been an important research direction in NLP for a long time. Since sequence-to-sequence model [1] have been applied to conversation generation, people began to pay attention to the research on open-domain dialogue system. In the early researches, researchers hope that the responses generated by dialogue model can be more meaningful, so many modes are proposed to improve the quality of generated responses, such as, using context information, promoting the diversity of the generated responses.

Compared with other directions, there are few researches focusing on the emotional factors in the dialogue system. The main reason is that there are some difficulties in this field. First, emotions in dialogue are changeable, and it is not easy to predict. Second, the lack of large-scale emotional dialogue datasets also makes difficult to do research. Finally, how to combine emotional information with text information is also a big challenge.

Most of the current researches on emotional dialogue system focus on how to integrate emotional information into single round dialogue. Models will generate a response with specified emotion, rather than learn and choose which emotion should be used to generate responses.

Our paper proposes a model to study the emotion in multi-turn dialogue, so that it can choose which emotion should be used to generate responses. As a result, the process of generating response will be closer to human’s dialogue.

2. Related Work
Emotion is an important aspect in conversation. Human have a lot of kinds of emotions, such as happiness, sadness, jealousy and so on. If we want to create a chatbot that can communicate with person, it is necessary to integrate emotion into the dialogue system. Earlier studies have proposed that
put emotional information into dialogue system can improve user’s satisfaction and the quality of response. However, the early methods are limited to small-scaled dataset. Until recently, emotional information has been applied to large-scale text generation and dialogue system.

Ghosh et al. [2] first proposed to integrate emotional factors into text generation. They proposed Affect-LM to add the required emotional information to the generated text. Hu et al. [3] proposed a model, which can generate sentences according to some properties of the text (emotion or tense, etc.).

After that, emotional text generation began to be applied to dialogue system. Huang et al. [4] first proposed to input the emotional word vector into the encoder or decoder part of the model, to add the required emotion to the response. Zhou et al. [5] propose Emotional Chatting Machine, which can response with required emotion. ECM is based on the sequence-to-sequence structure, and it has emotional embedding, internal memory, external memory and other modules to control the generation of emotional responses. Asghar et al. [6] integrated emotion into the dialogue model from three aspects: input (splicing emotion vector after word vector), training (emotional loss function) and inference (emotional beam search). Sun et al. [7] proposed a method based on Seq-GAN with emotion vector. Song et al. [8] proposed EmoDS, which added emotion classifier and attention on the basis of ECM, so that the generated response will be more inclined to required emotion.

Our model is based on ECM model and add multi-turn dialogue emotion modeling module, so that the model can learn which emotion should be used in the response. This makes up for the shortage of ECM that researchers need to specify emotion of the response.

3. Model

3.1. Background: Emotional Chatting Machine

Emotional Chatting Machine is based on the encoder-decoder structure, and includes four additional modules: emotion classifier, emotion embedding, internal memory and external memory. Emotion classifier is used to construct emotional dataset. Emotion embedding is to express emotion with vector. The internal memory module is used to simulate the dynamic changes of emotions. There is an internal emotion state for each category before the decoding process starts; at each step the emotion state decays by a certain amount and decay to zero when the decoding process is completed. The emotion update is completed through read gate and write gate.

Calculating read gate and write gate:

\[ g^r_t = \text{sigmoid}(W^r_e [e(y_{t-1}); s_{t-1}; c_t]) \]
\[ g^w_t = \text{sigmoid}(W^w_e s_t) \]

Updating emotion state:

\[ M^r_{t,t} = g^r_t \otimes M^r_{t,t} \]
\[ M^w_{t,t+1} = g^w_t \otimes M^w_{t,t} \]

The external memory module is used to control the expression of the model by assigning generation probability of common words and affective words. The type selector \( \alpha_t \) controls the weight of generating an emotion or a generic word.

Because ECM can achieve a nice balance between emotion expression and grammatical fluency, so our paper use ECM as the basic model.

3.2. Overview

The purpose of our work is to model the emotion in multiple rounds of conversation. Giving an input \( X = (x_1, ..., x_n) \), we need to get the emotion of output \( Y \). Then put input \( X \) and the emotion of output \( Y \) into ECM model, at last, we can get output \( Y = (y_1, ..., y_n) \).

It is a multi-classification problem to get the emotion of output \( Y \) from input \( X \) in multi-turn dialogue. Because LSTM shows great performance and it can capture long sequence relations, we choose LSTM model to predict the emotion of output \( Y \).
The input X is a sentence and should be mapped to a vector. We need to convert the words in X into vectors at first. Then we should map the word vectors to a sentence vector. This step can be achieved by LSTM, we put the words in the sentence X into LSTM in order, and takes state vector of last hidden layer as the sentence vector.

For multi-turn dialogue, we take the word vector of each word in 2N-1th sentence as input of encoder, and the state vector of last hidden layer will be put into a full connection layer then use SoftMax to get the emotion category of 2Nth sentence.

An overview of our model is given in Figure 1.

4. Dataset Preparation

The use of sections to divide the text of the paper is optional and left as a decision for the author. Due to there is few large-scale emotional dialogue datasets, Zhou et al. use emotional classifier to label NLPCC dataset. Although they use Bi-LSTM as emotion classifier, the accuracy of emotion classify is still only 0.623. It means that the dataset constructed by emotion classifier has some errors. Moreover, the NLPCC dataset is a single-turn dialogue dataset, which is not applicable to the multi-turn dialogue study.

By reading papers, we found that there exist multi-turn dialog dataset with emotional labels, such as IEMOCAP, SEMAINE, Emotionlines, MELD, DailyDialog and EmoContext. IEMOCAP, Emotionlines and MELD are multimoding (including audio, visual and text information), DailyDialog and EmoContext are text dataset. Table 1 shows the distribution of emotional tags for these datasets. We finally chose the DailyDialog as our dataset.

5. Experiments

5.1. Implementation Details

We use Tensorflow to implement the model. The encoder and decoder in ECM model use 2-layers Bi-LSTM with 256 hidden cells. The number of emotion categories is 7 and the dimension of emotion vector is 32. The dimension of word vector is 256 and vocabulary size is 40000. LSTM model use single layer LSTM, hidden size and the dimension of word vector are both 128.

| Label                  | DailyDialog | MELD | EmotionLines | IEMOCAP | EmoContext |
|------------------------|-------------|------|--------------|---------|------------|
| Neutral                | 85572       | 6436 | 6530         | 1708    | -          |
| Happiness/Joy          | 12885       | 2308 | 1710         | 648     | 4669       |
| Surprise               | 1823        | 1636 | 1658         | -       | -          |
| Sadness                | 1150        | 1002 | 498          | 1084    | 5838       |
| Anger                  | 1022        | 1607 | 772          | 1103    | 5954       |
5.2. Baseline
Although our work is based on ECM, due to our research direction is to let the model learn emotion in multi-turn dialogue, however, ECM generates response with required emotion. As a result, our work can not be directly compared with ECM, so we choose the basic sequence-to-sequence model as the baseline.

5.3. Evaluation
First of all, we need to choose which model to use in modeling emotions in multi-turn dialogue. The essence of emotion modeling in multi-turn dialogue is similar to the emotion classification problem. So, we compared the emotion-classification performance of each model on DailyDialog dataset. Table 2. shows the performance of each model. Finally, we choose LSTM to capture emotions in multi-turn dialogue because it has best performance.

Table 2. Emotion-classification of each model on DailyDialog.

| Model     | Accuracy |
|-----------|----------|
| RNN       | 0.575    |
| CNN       | 0.583    |
| LSTM      | 0.668    |
| Bi-LSTM   | 0.635    |

We choose perplexity and accuracy to evaluate the performance of the model. Perplexity is used to evaluate sentence from grammar and fluency. Accuracy is used to evaluate the consistency between required emotion and the emotion of generated sentence. Table 3. shows the evaluation result of sequence-to-sequence model and our model.

Table 3. Evaluation with perplexity and accuracy

| Method    | Perplexity | Accuracy |
|-----------|------------|----------|
| Seq2Seq   | 67.5       | 0.192    |
| Our model (LSTM) | 66.4      | 0.619    |

The result shows that, compared with the sequence-to-sequence model, our model performs better from perplexity and accuracy two aspects. It means that our model has a better performance in emotion generation and sentence fluency than sequence-to-sequence model.

Compared with ECM model, our model can learn the law of emotional change in multi-turn dialogue by itself. And in the infer section, we do not need to input the required emotion of output sentence. The model can determine which emotion type should be used in the output sentence, which is much more similar to human conversation.

6. Conclusion
In this paper, we propose a model to learn the emotional change in multi-turn dialogue, so that researchers do not need to input emotion during the infer section. And we used DailyDialog as our dataset instead of using emotion classifier to construct the dataset, which improves accuracy of the dataset. Compared with traditional sequence-to-sequence model, our model performs better in emotion generation and sentence fluency.

In our future work, we will try to use context information in our model and improve accuracy of emotion predict.
References

[1] Sutskever, I., Vinyals, O., and Le, Q. V. (2014) Sequence to sequence learning with neural networks. Advances in neural information processing systems 27: 3104-3112.

[2] Ghosh, S., Chollet, M., Laksana, E., Morency, L. P., and Scherer, S. (2017). Affect-lm: A neural language model for customizable affective text generation. arXiv preprint arXiv:1704.06851.

[3] Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R., and Xing, E. P. (2017). Toward controlled generation of text. arXiv preprint arXiv:1703.00955.

[4] Huang, C., Zaiane, O. R., Trabelsi, A., and Dziri, N. (2018). Automatic dialogue generation with expressed emotions. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers) (pp. 49-54).

[5] Zhou, H., Huang, M., Zhang, T., Zhu, X., and Liu, B. (2017). Emotional chatting machine: Emotional conversation generation with internal and external memory. arXiv preprint arXiv:1704.01074.

[6] Asghar, N., Poupart, P., Hoey, J., Jiang, X., and Mou, L. (2018). Affective neural response generation. In European Conference on Information Retrieval (pp. 154-166). Springer, Cham.

[7] Sun, X., Chen, X., Pei, Z., & Ren, F. (2018, May). Emotional human machine conversation generation based on SeqGAN. In 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia) (pp. 1-6). IEEE.

[8] Song, Z., Zheng, X., Liu, L., Xu, M., and Huang, X. J. (2019, July). Generating responses with a specific emotion in dialog. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3685-3695).