Learning to Converse Emotionally Like Humans: A Conditional Variational Approach

Rui Zhang and Zhenyu Wang(✉)

Department of Software Engineering, South China University of Technology, Guangzhou, People's Republic of China
z.rui16@mail.scut.edu.cn, wangzy@scut.edu.cn

Abstract. Emotional intelligence is one of the key parts of human intelligence. Exploring how to endow conversation models with emotional intelligence is a recent research hotspot. Although several emotional conversation approaches have been introduced, none of these methods were able to decide an appropriate emotion category for the response. We propose a new neural conversation model which is able to produce reasonable emotion interaction and generate emotional expressions. Experiments show that our proposed approaches can generate appropriate emotion and yield significant improvements over the baseline methods in emotional conversation.

Keywords: Emotion selection · Emotional conversation

1 Introduction

The ability of a computer to converse in a natural and coherent manner with humans has long been held as one of the primary objectives of artificial intelligence, yet conventional dialog systems continue to face challenges in emotion understanding and expression.

In the past, the research on emotional response generation focused on the domain-specific, task-oriented dialogue systems. These methods are mainly based on some hand-crafted emotion inference rules to choose a reasonable strategy, and retrieve a suitable pre-designed template for response generation.

However, these approaches are not suitable for open-domain chatterbot, since there are large amount of topics and more complex emotion states in chitchat conversation. For instance, if the user says "my cat died yesterday", it is reasonable to generate response like "so sorry to hear that" to express sadness, also it is appropriate to generate response like "bad things always happen, I hope you will be happy soon" to comfort the user. Although some neural network based methods for open-domain emotional conversation [3,17,18,20] have been proposed recently, yet these approaches mainly focused in generating emotional

M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11108, pp. 98–109, 2018.
https://doi.org/10.1007/978-3-319-99495-6_9

expressions and none of them provide a mechanism to determine which emotion category is appropriate for response generation.

This paper presents two kinds of emotion sensitive conditional variational autoencoder (EsCVAE) structure for determining a reasonable emotion category for response generation. Just the same as children learn to converse emotionally through imitation, the principle idea is to model the emotion interaction patterns from large-scale dialogue corpus as some kind of distribution, and sample from this emotion interaction distribution to decide what kind of emotion should be expressed when generating responses.

As far as we know, this is the first work to determine an appropriate emotion category for current large-scale conversation generation model. To sum up, our contributions are as follows:

- We propose the EsCVAE-I and EsCVAE-II model to learn to automatically specify a reasonable emotion category for response generation. Experiments show that our proposed approaches yield significant improvements over the baseline methods.
- We show that there are some frequent emotion interaction patterns in humans dialogue (e.g. happiness-like, angry-disgust), and our models are able to learn such frequent patterns and apply it to emotional conversation generation.

2 Related Work

2.1 Conversation Generation

In recent years, there is a surge of research interest in dialogue system. Due to the development of deep neural network, learning a response generation model within a machine translation (MT) framework from large-scale social conversation corpus becomes possible. Following the principle idea of sequence-to-sequence architecture, recurrent network based models [11,16] and VAE based models [2,10] were successively proposed. The basic idea of VAE is to firstly encode the input x into a probability distribution z, and then to apply a decoder network to reconstruct the original input x using samples from z.

To better control the generative process, the conditional variational autoencoder (CVAE) [4,12] is recently introduced to generate diverse texts conditioned on certain attributes c. The conditional distribution in the CVAE is defined as $p(x, z|c) = p(x|z, c)p(z|c)$. By approximating $p(x|z, c)$ and $p(z|c)$ using deep neural network (parameterized by θ), the generative process of x is: (1) sample a latent variable z from the prior network $p_\theta(z|c)$, and (2) generate x through the response decoder $p_\theta(x|z, c)$. As proposed in [15], CVAE can be trained by maximizing the variational lower bound of the conditional log likelihood. By assuming the z follows Gaussian distribution and introducing a recognition network $q_\phi(z|x, c)$ to approximate the true posterior distribution $p(z|x, c)$, the variational lower bound can be written as:

$$\begin{aligned}\mathcal{L}(\theta, \phi; x, c) &= -KL(q_\phi(z|x, c)||p_\theta(z|c)) + \mathbf{E}_{q_\phi(z|x,c)}[\log p_\theta(x|z, c)] \\ &\leq \log p(x|c)\end{aligned} \tag{1}$$

Further, Zhao et al. [19] introduced Knowledge-Guided CVAE (kgCVAE) to get benefits from linguistic cues. They assumed that in kgCVAE the generation of x also depends on the linguistic cues y, and have proved that the kgCVAE based dialogue model can more easily control the model's output by cooperating with the linguistic cues.

2.2 Emotional Intelligence

Emotional intelligence is one of the key parts of human intelligence. Exploring the influence of emotional intelligence on human-computer interaction has a long history. Experiments show that dialogue systems with emotional intelligence lead to less breakdowns in dialogue [6], and enhance users' satisfaction [9].

In the early studies, a few of emotion modeling approaches were introduced to construct emotional dialogue systems. Polzin et al. [8] proposed a pioneer work in emotional human-computer conversation, which is capable of employing different discourse strategies based on users' affection states. Andre et al. [1] integrates social theory of politeness with cognitive theory of emotions to endow dialogue systems with emotional intelligence. Skowron [13,14] quantize users' emotion via affective profile and respond to users' utterances at content- and affect- related levels.

Recently, neural network based methods for emotional text generation have been investigated. Zhou et al. [20] proposed Emotional Chatting Machine (ECM) to generate emotional response by adopting emotion category embedding, internal emotional memory and external memory. Ghosh et al. [3] introduced Affect-LM which is able to generate emotionally colored conversational text in five specific affect categories with varying affect strengths.

Unfortunately, none of these neural network based approaches provides a mechanism to determine which emotion category is appropriate for emotional response generation. These works, mainly focused in generating emotional expressions, are either driven by pre-defined rules, or in need of specifying target emotion manually. Thus, in this paper we introduce two kinds of EsCVAE model to address this problem.

3 Proposed Models

3.1 Problem Definition

Our models aim to learn the inner relationship of emotional interaction, and to automatically specify a reasonable emotion category for response generation given an input utterance. In practice, however, it is hard to evaluate the appropriateness if we only predict the emotional category. Therefore, we reformulate the task as below:

Given a post utterance $\mathbf{u}_p = (w_{p_1}, w_{p_2}, \cdots, w_{p_m})$ and its emotion category $\mathbf{e}_p$, the goal is to predict a reasonable emotion category $\mathbf{e}_r$ for response generation, and generate a response utterance $\mathbf{u}_r = (w_{r_1}, w_{r_2}, \cdots, w_{r_n})$ which is coherent

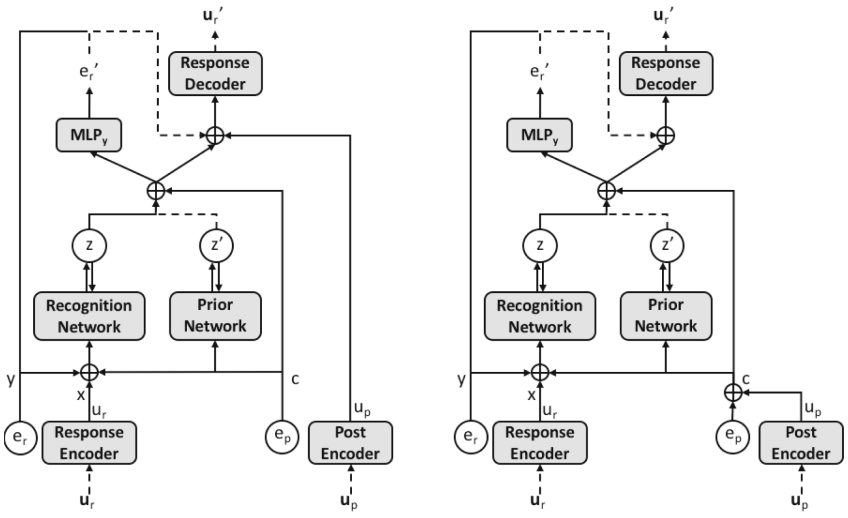
with $\mathbf{e}_r$, where w_{p_k} and w_{r_k} are the k-th words in the post utterance and the response utterance, respectively. The emotions are divided into six categories {*Anger, Disgust, Happiness, Like, Sadness, Other*}.

3.2 EsCVAE-I: Conditioned on Emotions only

Since most rule-based strategies are triggered according to the emotion type of the input text, we firstly consider a variational autoencoder architecture conditioned on emotional information.

Figure 1(a) delineates an overview of our EsCVAE-I model. An emotional category embedding network is adopted to represent the emotion category of the utterance by a real-value, low dimensional vector, since an emotion category provides a high-level abstraction of an expression of the emotion. We randomly initialize the vector of each emotion categories, and then learn the vectors of emotion category through training. Thus, the emotion categories of the post utterance and the response utterance are represented by emotion embedding vectors e_p and e_r, respectively. A bidirectional GRU network is adopted as the response encoder to encode the response utterance $\mathbf{u}_r$ into a fixed-sized vector u_r, by concatenating the last hidden states of the forward and backward RNN. The post encoder is another GRU network that encodes the post utterance $\mathbf{u}_p$ into a vector u_p.

To capture the inner relationship of emotional interaction, a conditional variational autoencoder architecture is applied in the proposed model. In EsCVAE-I model we consider the post emotion e_p as the condition c and response emotion e_r as the linguistic feature y. The target utterance x is simply the response



(a) EsCVAE-I (b) EsCVAE-II

Fig. 1. Illustrations of our proposed models.

utterance u_r. Assuming latent variable z follows isotropic Gaussian distribution, with the recognition network $q_\phi(z|x,c,y) \sim \mathcal{N}(\mu, \sigma^2\mathbf{I})$ and the prior network $p_\theta(z|c) \sim \mathcal{N}(\mu', \sigma'^2\mathbf{I})$, we have:

$$\begin{bmatrix} \mu \\ \log(\sigma^2) \end{bmatrix} = W_r \begin{bmatrix} x \\ c \\ y \end{bmatrix} + b_r = W_r \begin{bmatrix} u_r \\ e_p \\ e_r \end{bmatrix} + b_r \tag{2}$$

$$\begin{bmatrix} \mu' \\ \log(\sigma'^2) \end{bmatrix} = \text{MLP}_p(c) = W_p e_p + b_p \tag{3}$$

In the training stage, we obtain samples of z from $\mathcal{N}(z; \mu, \sigma^2\mathbf{I})$ predicted by the recognition network. The response decoder is a GRU network with initial state $s_0 = W_i[z, c, e_r, u_p] + b_i$, which then predicts the words in x sequentially. An MLP is adopted to predict the response emotion category $e'_r = \text{MLP}_y(z, c)$ based on z and c. While testing, samples of z is obtained from $\mathcal{N}(z; \mu', \sigma'^2\mathbf{I})$ predicted by the prior network. And the initial state of the response decoder is calculated as: $s_0 = W_i[z, c, e'_r, u_p] + b_i$, where e'_r is the predicted response emotion.

The proposed model is trained by minimizing the reconstruction loss while maximizing the variational lower bound. The reconstruction loss is calculated based on the cross entropy error. Following [19] the variational lower bound can be calculated as:

$$\begin{aligned} \mathcal{L}(\theta, \phi; x, c, y) = &-KL(q_\phi(z|x,c,y)||p_\theta(z|c)) \\ &+ \mathbf{E}_{q_\phi(z|x,c,y)}[\log p(x|z,c,y)] \\ &+ \mathbf{E}_{q_\phi(z|x,c,y)}[\log p(y|z,c)] \end{aligned} \tag{4}$$

3.3 EsCVAE-II: Sensitive to both Content-Level and Emotion-Level Information

A natural extension of the previous approach is a model that is also sensitive to the content information, since emotion interactions in dialogue are not only related to the emotional state of the talkers, but also closely related to the topic of the conversation.

In order to get benefits from these features, we propose the EsCVAE-II model. In this model we consider both the content-level and emotion-level information of the post utterance as condition c, by concatenating the utterance and emotion vectors: $c = [u_p, e_p]$. Following the same assumption, the recognition network $q_\phi(z|x,c,y)$ and the prior network $p_\theta(z|c)$ are calculated as:

$$\begin{bmatrix} \mu \\ \log(\sigma^2) \end{bmatrix} = W_r \begin{bmatrix} x \\ c \\ y \end{bmatrix} + b_r = W_r \begin{bmatrix} u_r \\ [u_p, e_p] \\ e_r \end{bmatrix} + b_r \tag{5}$$

$$\begin{bmatrix} \mu' \\ \log(\sigma'^2) \end{bmatrix} = \text{MLP}_p(c) = W_p[u_p, e_p] + b_p \tag{6}$$

Then we obtain samples of z either from $\mathcal{N}(z; \mu, \sigma^2\mathbf{I})$ predicted by the recognition network (while training) or $\mathcal{N}(z; \mu', \sigma'^2\mathbf{I})$ predicted by the prior network (while testing). Since the content information of post utterance u_p has been contained in c, the initial state of the response decoder is therefore changed to $s_0 = W_i[z, c, e_r] + b_i$. Details of the EsCVAE-II model are shown in Fig. 1(b).

4 Experiment Setup

4.1 Datasets

The NLPCC Corpus We use the Emotional Conversation Dataset of NLPCC 2017, which consists of 1,119,207 post-response pairs, to evaluate our approaches. The dataset is automatically annotated by a six-way Bi-LSTM classifier which is reported to reach an accuracy of 64%[1].

The STC-2 Corpus We also use the Short Text Conversation Corpus (STC-2) dataset, which consists of 4,433,949 post-response pairs collected from Weibo. We apply the same Bi-LSTM classifier to annotate this corpus.

4.2 Model Details

We implement the encoders as 2-layer GRU and the response generator as single-layer GRU RNNs, with input and hidden dimension of 400 and maximum utterance length of 25. The dimensions of word embedding and emotion embedding are also set to 400. We use a KL term weight linearly annealing from 0 to 1 during training, to avoid vanishingly small KL term in the VAE module as introduced in [2]. Both the prior network and the recognition network consist of 200 hidden neurons.

In order to generate a better response, we adopt the learning to start (LTS) technique [21], which use an additional network layer to predict the first word instead of using a "GO" symbol as the first word. In addition, beam-search is also adopted with beam size of 5.

4.3 Evaluation Results

Since automatically evaluating an open-domain generative dialog model is still an open research challenge [5], we provide the following metrics to measure our models.

Quantitative Analysis The following metrics are proposed to automatically evaluate our models. At the content level, perplexity and BLEU are adopted. At the emotion level, we introduce emotion accuracy and EIP difference degree as the evaluating metrics. We randomly sample 5000 posts for test. Details of quantitative analysis results are reported in Table 1.

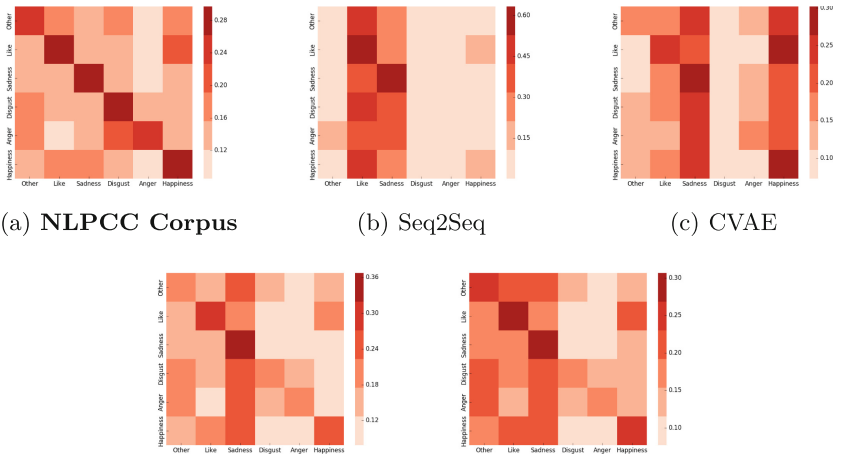
[1] http://tcci.ccf.org.cn/conference/2017/dldoc/taskgline04.pdf.

Table 1. Performance of each model on automatic measures. Note that our BLEU scores has been normalized to [0,1].

Corpora	Model	Perplexity	BLEU	Acc.	Diff.
NLPCC corpus	Seq2Seq	101.0	0.105	-	1.195
	CVAE	47.2	0.090	-	0.171
	EsCVAE-I	46.1	0.114	0.675	0.085
	EsCVAE-II	**44.3**	**0.139**	**0.690**	**0.072**
STC-2 corpus	Seq2Seq	88.4	0.156	-	0.289
	CVAE	20.7	0.224	-	0.048
	EsCVAE-I	20.9	0.211	0.621	0.042
	EsCVAE-II	**19.0**	**0.230**	**0.627**	**0.035**

BLEU is a popular metric which measures the geometric mean of modified n-gram precision with a length penalty [7]. We use BLEU-3 as lexical similarity metrics and normalize the score to [0,1]. Following [16] we also employ perplexity as an evaluation metric at the context level.

We quantitatively measure emotion accuracy to evaluate the EsCVAE-I and EsCVAE-II, which is defined as the agreement between the expected emotion category e'_r (generated by our models) and emotion category of the corresponding generated response (predicted by the Bi-LSTM classifier mentioned in Sect. 4.1).



(a) **NLPCC Corpus** (b) Seq2Seq (c) CVAE

(d) EsCVAE-I (e) EsCVAE-II

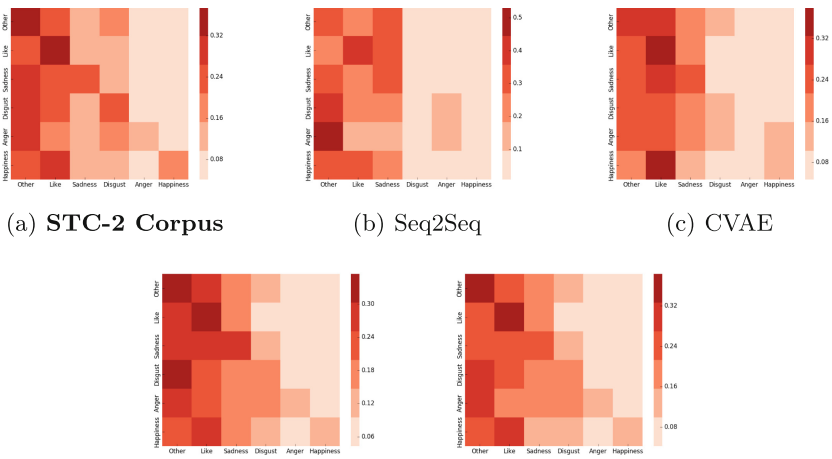
Fig. 2. Visualization of emotion interaction on NLPCC Corpus. The emotion categories on X/Y-axis (from left/top to right/bottom) are: *Other, Like, Sadness, Disgust, Anger, Happiness.*

To evaluate the imitative ability of the emotional dialogue model, we adopt the EIP difference degree as the evaluating metric. Emotion interaction pattern (EIP) is defined as the pair of emotion categories of a post and its response. The value of an EIP is the conditional probability $P(e_r|e_p) = P(e_r, e_p)/P(e_p)$. We define the EIP difference degree as the agreement between the EIP distribution of the original corpus and the EIP distribution of the generated results, calculated as:

$$\text{Difference degree} = \sum_{r} \sum_{p} |P_o(e_r|e_p) - P_g(e_r|e_p)|^2 \tag{7}$$

where $P_o(e_r|e_p)$ stands for the emotion interaction distribution of the generated results, while $P_g(e_r|e_p)$ is the distribution of the original corpora. In other words, a lower difference degree indicates that the model has a better ability to imitate emotion interaction.

As shown in Table 1, we have the following observations: (1) CVAE-based seq2seq model performs much better than vanilla seq2seq model, since vanilla seq2seq model tends to generate "safe" and dull responses, which lead to higher difference degree and higher perplexity, and (2) thanks to the external linguistic feature (response emotion), both EsCVAE models perform better on difference degree metric. Figures 2 and 3 visualize the EIP distribution by heat maps, where the color darkness indicates the strength of an interaction. We found that both EsCVAE models perform better on emotion interaction imitating.



(a) **STC-2 Corpus** (b) Seq2Seq (c) CVAE

(d) EsCVAE-I (e) EsCVAE-II

Fig. 3. Visualization of emotion interaction on STC-2 Corpus.

Human Evaluation In order to better evaluate the proposed model at both content- and emotion- level, we also analyze the generated results by human evaluation. We recruit 3 annotators for human evaluation experiments. Annotators are asked to score the generated responses in terms of emotional accuracy and naturalness. Naturalness are annotated according to the following criteria:

```
IF (Coherence and Fluency)
    IF (Emotion Appropriateness)
        SCORE 2
    ELSE
        SCORE 1
ELSE
    SCORE 0
```

We extract 300 posts from the test set, and for each model we generate 3 responses for each post. The results of human evaluation for the quality of response are shown in Table 2. We calculate the Fleiss' kappa as the statistical measure of inter-rater consistency, the average score of Fleiss' kappa for naturalness is 0.471 and for emotion accuracy is 0.744. As can be seen, the EsCVAE-I model gets the best performance in emotion accuracy metric, while the EsCVAE-II model achieves better score in naturalness metric.

Table 2. Human evaluation in terms of emotion accuracy and naturalness.

Corpora	Model	Acc.	Naturalness
NLPCC corpus	Seq2Seq	-	1.080
	CVAE	-	1.163
	EsCVAE-I	**0.283**	1.176
	EsCVAE-II	0.250	**1.217**
STC-2 corpus	Seq2Seq	-	1.170
	CVAE	-	1.187
	EsCVAE-I	**0.367**	1.193
	EsCVAE-II	0.353	**1.250**

In the progress of human evaluation, we found that vanilla seq2seq model failed to capture emotion interaction, and most dull responses generated by this model, such as "Me too" and "Haha", are considered as emotionally inappropriate. On the other hand, the EsCVAE-I and EsCVAE-II model are capable of learning frequent EIPs from the training corpus and generating reasonable emotional responses. For example, when the post utterance is in a negative emotional state (anger, disgust, ...), the EsCVAE based models tend to predict the negative emotional category, and vice versa (as shown in Table 3).

In addition, the emotion accuracy in human evaluation is much lower than in automatically evaluation, which indicates that it is still hard to assess a dialogue model at emotion level, since this metric is complicate and subjective.

Table 3. Sample responses generated by each model (original Chinese and English translation).

Post	Model		Response(Original)	Response(Translated)
我老了，已经不会再写诗了。能做的就是让你在我的怀里躲风。 I am old and I can no longer write poetry. What I can do is hide you in my arms. **(Sadness)**	Seq2Seq		我想知道你是怎么做到的。	I want to know how you did it.
	CVAE	Like	好可爱啊	So cute.
		Happiness	哈哈，诗人，原来是这样	Haha, the poet, it turned out to be this way.
	EsCVAE-1	Sadness	你真的很伤感啊！	You are really sad!
		Sadness	你看不见，你永远不懂	You will never understand what you cannot see.
	EsCVAE-2	Like	我也喜欢你的书……	I like your book too.
		Disgust	你的梦还没醒呢？	Didn't you wake up from your dreams?
将来的你，一定会感谢现在拼命努力的你。 In the future, you will certainly thank you for your hard work. **(Like)**	Seq2Seq		我现在也是这样的。	I am like this now.
	CVAE	Other	你知道将来的我吗？	Do you know me in the future?
		Happiness	是的，这让我们在一起！	Yes, this allows us to be together!
	EsCVAE-1	Like	我的青春也不能虚度啊。	My youth cannot be devalued.
		Other	将来的你，你会后悔。	In the future, you will regret it.
	EsCVAE-2	Like	嗯嗯嗯，绝对的。	Um, absolutely.
		Like	坚持就会有收获的。	As long as you persist, there will be a harvest.

5 Conclusion

In this paper, we propose two kinds of emotion sensitive conditional variational autoencoder (EsCVAE) structure for emotional conversation, to model the inner relationship of emotional interaction in human dialogue and generate emotional responses. To our best knowledge, this is the first work to generate an appropriate emotion category for current large-scale conversation generation model. Automatic and manual evaluation results show that EsCVAE based models can predict a reasonable emotion category for response generation by learning emotion interaction pattern from the training corpus.

We leave the exploration of EsCVAEs with attention mechanism for future work. Additional accuracy improvements might be also achieved by extended features (e.g. topics, dialog-act). At the same time, we will improve our model by considering polite rules and persona model to avoid generating offensive responses. We also plan to investigate the applicability of our model for task-oriented conversation.

Acknowledgements. This work is supported by the Science and Technology Program of Guangzhou, China(No. 201802010025), the Fundamental Research Funds for the Central Universities(No. 2017BQ024), the Natural Science Foundation of Guangdong Province(No. 2017A030310428) and the University Innovation and Entrepreneurship Education Fund Project of Guangzhou(No. 2019PT103). The authors also thank the editors and reviewers for their constructive editing and reviewing, respectively.

References

1. André, E., Rehm, M., Minker, W., Bühler, D.: Endowing spoken language dialogue systems with emotional intelligence. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 178–187. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24842-2_17
2. Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A., Jozefowicz, R., Bengio, S.: Generating sentences from a continuous space. In: Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, pp. 10–21 (2016)
3. Ghosh, S., Chollet, M., Laksana, E., Morency, L.P., Scherer, S.: Affect-LM: a neural language model for customizable affective text generation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 634–642 (2017)
4. Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R., Xing, E.P.: Toward controlled generation of text. In: International Conference on Machine Learning, pp. 1587–1596 (2017)
5. Liu, C.W., Lowe, R., Serban, I., Noseworthy, M., Charlin, L., Pineau, J.: How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2122–2132 (2016)
6. Martinovski, B., Traum, D.: Breakdown in human-machine interaction: the error is the clue. In: Proceedings of the ISCA Tutorial and Research Workshop on Error Handling in Dialogue Systems, pp. 11–16 (2003)
7. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics (2002)
8. Polzin, T.S., Waibel, A.: Emotion-sensitive human-computer interfaces. In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion (2000)
9. Prendinger, H., Mori, J., Ishizuka, M.: Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. Int. J. Hum. Comput. Stud. **62**(2), 231–245 (2005)
10. Semeniuta, S., Severyn, A., Barth, E.: A hybrid convolutional variational autoencoder for text generation. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 627–637 (2017)
11. Shang, L., Lu, Z., Li, H.: Neural responding machine for short-text conversation. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), vol. 1, pp. 1577–1586 (2015)
12. Shen, X., Su, H., Li, Y., Li, W., Niu, S., Zhao, Y., Aizawa, A., Long, G.: A conditional variational framework for dialog generation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, pp. 504–509 (2017)

13. Skowron, M.: Affect listeners: acquisition of affective states by means of conversational systems. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) Development of Multimodal Interfaces: Active Listening and Synchrony. LNCS, vol. 5967, pp. 169–181. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12397-9_14
14. Skowron, M., Rank, S., Theunis, M., Sienkiewicz, J.: The good, the bad and the neutral: affective profile in dialog system-user communication. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011. LNCS, vol. 6974, pp. 337–346. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24600-5_37
15. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: Advances in Neural Information Processing Systems, pp. 3483–3491 (2015)
16. Vinyals, O., Le, Q.: A neural conversational model. arXiv preprint arXiv:1506.05869 (2015)
17. Yuan, J., Zhao, H., Zhao, Y., Cong, D., Qin, B., Liu, T.: Babbling - The HIT-SCIR system for emotional conversation generation. In: Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Y. (eds.) NLPCC 2017. LNCS (LNAI), vol. 10619, pp. 632–641. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73618-1_53
18. Zhang, R., Wang, Z., Mai, D.: Building emotional conversation systems using multi-task Seq2Seq learning. In: Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Y. (eds.) NLPCC 2017. LNCS (LNAI), vol. 10619, pp. 612–621. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73618-1_51
19. Zhao, T., Zhao, R., Eskenazi, M.: Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 654–664 (2017)
20. Zhou, H., Huang, M., Zhang, T., Zhu, X., Liu, B.: Emotional chatting machine: emotional conversation generation with internal and external memory. arXiv preprint arXiv:1704.01074 (2017)
21. Zhu, Q., Zhang, W., Zhou, L., Liu, T.: Learning to start for sequence to sequence architecture. arXiv preprint arXiv:1608.05554 (2016)