ELEKTROTEHNIŠKI VESTNIK 87(3): 127-134, 2020
ORIGINAL SCIENTIFIC PAPER

A Sentimental and Context-Sensitive Model for the Seq2Seq-Based Dialogue Generation

Zhiqiang Ma[1,2], Baoxiang Du[1], Ji Shen[1], Chunyu Wang[1], Rui Yang[1]

[1] *College of Data Science and Application, Inner Mongolia University of Technology, China*
[2] *Inner Mongolia Autonomous Region Engineering & Technology Research Centre of Big Data Based Software Service, China*
E-mail: mzq_bim@imut.edu.cn

Abstract. For the dialogue generation task, the Sequence to Sequence (Seq2Seq) model lacks presentation of the sentimental information and dialogue context in the encoding procedure, resulting in generating sentimentally poor and context-irrelevant responses. To address these problems, the paper proposes a sentimental and context-sensitive Seq2Seq model for dialogue generation. For the Seq2Seq model to contain sentiments, a sentimental word vector and utterance sentiment vector are built to reinforce the sentimental information presentation. To make the generated responses contextually more relevant, an utterance encoder and context encoder are used. Experiments show that the proposed model generates more sentimental, contextually-relevant and high-quality responses.

Keywords: dialogue generation; Sequence to Sequence (Seq2Seq); sentimental information representation; context-sensitive model

Model za generacijo dialoga z upoštevanjem čustev in vsebine sporočila

Pri generaciji dialoga model zaporedje v zaporedje (Seq2Seq) ni natančen pri upoštevanju čustev, česar posledica so nepravilni in vsebinsko nepopolni odzivi. V članku predlagamo model Seq2Seq, ki bo upošteval tudi čustva in vsebino sporočila pri generaciji dialoga. Za upoštevanje čustvenih informacij smo zgradili čustveni besedni vektor.
Z namenom, da so ustvarjeni odzivi vsebinsko smiselni, smo zgradili govorni in vsebinski kodirnik. Eksperimentalni rezultati so potrdili, da predlagani model generira boljše dialoge.

1 INTRODUCTION

The dialogue model is considered to be one of the highest goals of artificial intelligence, and the earliest research of the dialogue models can be traced back to the Turing Test [1]. According to the core technology, the dialogue models can be categorized as artificial-templates-based dialogue model (e.g., Weizenbaum et al. [2]), retrieval-based dialogue model (e.g., Inaba et al. [3]), and generative dialogue model (e.g., Shang et al. [4]). In recent years, the rapid development of the deep neural networks has greatly promoted the research of the generative dialogue models. Ritter et al. [5] were the first to realize the dialogue generation system by using the idea of statistical machine translation. Vinyals et al. [6] were the first to apply the Sequence to Sequence (Seq2Seq) model in the dialogue generation system and obtained effective experimental results in the natural language dialogue. Based on the Seq2Seq model, researchers carried out the context-sensitive Seq2Seq model (e.g., Sordoni et al. [7]), allowing the model to consider the context information. The response diversity Seq2Seq model (e.g., Li et al. [8] and Mou et al. [9]) addressed the commonplace generation problem (e.g., the high occurrence of "I don't know"), researchers added a maximum mutual information as the objective function or used a content-introducing method to reduce the commonplace generation probability. The personalized Seq2Seq model (e.g., Li et al. [10]) captured the background and speaking style to maintain the speaker consistency in neural response generation. The Seq2Seq model has become the most widely used generative dialogue model.

To the best of our knowledge, we find the researches of the generative dialogue model have been mainly focused on the grammaticality, diversity and topic relevance of sentences, and there have been that few studies on sentiment generation in dialogue. Ghosh et al. [11] were the first to introduce the emotional information in the language model (in 2017), enabling the language model to generate an emotionally colored conversational text. However, sentiment is a necessary factor for the semantic expression in human dialogue. Sentimental representation and generation ability seriously affect the application of the dialogue generation model. The Seq2Seq model does not specifically consider sentimental factors in the utterances input, which makes the model lack the

Received 10 February 2020
Accepted 23 April 2020

sentimental information representation, resulting in the inability to stably generate sentimental dialogue responses. At the same time, the context information of the dialogue is not considered in the Seq2Seq model, especially the sentimental information in the context, which leads to the inability to ensure the sentimental consistency in the generated responses. Therefore, the Seq2Seq model is mainly limited by the sentimental dialogue sentences in the corpus when generating sentimental responses, and cannot stably generate sentimental responses consistent with the contextual semantics.

Following the above, in this paper we propose a sentimental and context-sensitive Seq2Seq model for dialogue generation. Our key contributions are two-fold: (1) our encoding of the utterances in the model adds a word-level and utterance-level sentimental information, which reinforces the model representation ability of sentiments; (2) the encoder generates the utterance and context vectors containing an utterance-level and context-level semantic information, respectively, so that the decoder generates both the sentimental and contextual dialogue responses.

2 Related Work

Seq2Seq was proposed by Sutskever et al. [12] and Cho et al. [13]. It consists of an encoder, hidden state vector and decoder. The encoder generates the hidden state vector from the source sequence, and then the hidden state vector acts as an input to the decoder generating a target sequence for the output. The encoder and decoder are usually composed of Recurrent Neural Network (RNN) [14] units. RNN is a time steps model that records the above information mainly through the accumulation of time. However, RNN easily causes the vanishing gradient problem which disables the neural network to update the parameter and the long-distance dependence disappears. At the same time, RNN is also prone to cause the exploding gradient problem, which makes the network difficult to converge or even not converge. Therefore, based on RNN, a Long Short-Term Memory (LSTM) [15] model is proposed to solve the problem of RNN using the input gate, output gate, forget gate and cell state. However, the LSTM calculation is relatively complicated. Then, based on LSTM, the Gated Recurrent Unit (GRU) [16] model to simplify the gate calculation is proposed.

Zhou et al. [17] proposed an Emotional Chatting Machine (ECM), which is the first generative chatbot with emotion. For the first time, ECM adds an emotional information to the Seq2Seq-based dialogue generation model, which mainly uses a static emotion vector embedding representation, dynamic emotion states memory network and emotional word external memory mechanism to generate emotional responses according to the user input and specified emotional categories.

Asghar et al. [18] proposed a Seq2Seq-based dialogue model for an Affective Neural Response Generation (ANRG) which adds three-dimensional emotion vectors to the word vector, designs different loss functions to achieve smoothing and solves the minimizing and maximizing effective dissonance and other issues. ANRG improves the open-domain conversational prowess of the encoder-decoder networks by enabling them to produce emotionally rich responses that are more interesting and natural.

Our model improves the encoding procedure of the Seq2Seq-based dialogue generation model, by reinforcing representation of the sentimental information and dependence of the contextual information, thus enabling the model to generate sentimental and contextually relevant dialogue responses.

3 Sentimental and Context-Sensitive Seq2Seq Model

In order to generate sentimental and contextually relevant responses for the dialogue generation task, the sentimental information (word-level and utterance-level sentiment) and the contextual information must be comprehensively considered. Therefore, based on the Seq2Seq model, we designed a sentimental and context-sensitive dialogue generation model named the SCS-Seq2Seq model (see Fig. 1).

Our model generates the corresponding sentimental word vectors and utterance sentiment vectors according to the user current input (U_c) (the utterance of the

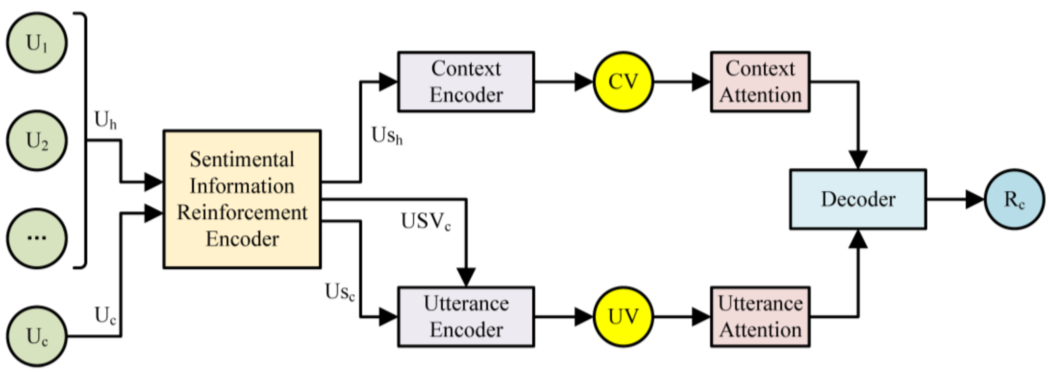

Figure 1. Architecture of the SCS-Seq2Seq model.

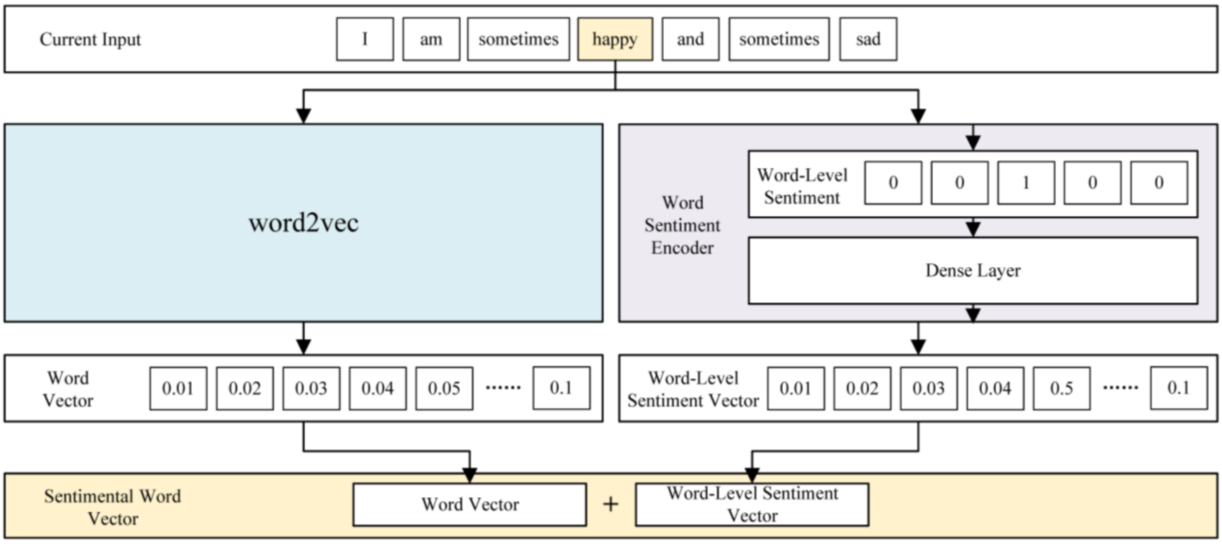

Figure 2. Process of the sentimental word encoding.

current turn of the dialogue) and the context input (U_h) (the context of the previous utterances of the dialogue) through the sentimental information reinforcement encoder. The utterance encoder inputs the sentimental word vector sequence (Us_c) and utterance sentiment vector (USV_c) corresponding to the current dialogue utterance, and encodes the utterance vector (UV); The context encoder inputs the sentimental word vector sequences (Us_h) corresponding to the dialogue context, and encodes the context vector (CV). The decoder generates a predicted dialogue response (R_c) according to the utterance vector (UV) and the context vector (CV) with an attention mechanism.

3.1 *Sentimental Information Reinforcement Encoding*

The key to the sentimental dialogue response generation is a sentimental information reinforcement encoder. This section focuses on the sentimental information reinforcement encoder for the Seq2Seq-based dialogue generation model, including the sentimental word encoding and the sentence sentiment encoding mechanism.

3.1.1 *Sentimental Word Encoding*

Giving an utterance "I am sometimes happy and sometimes sad", the sentimental words "happy" and "sad" in the utterance can be expressed by the one-hot method. "Happy" [0,0,0,1,0,0,0] is in the 4th position of the utterance, and "sad" [0,0,0,0,0,0, -1] is in the 7th position of the utterance. However, because the length of the utterances is inconsistent in practice, the method only describes the relative position information of the sentimental words and loses the absolute position information. We use the word sentiment encoder to fix the word-level sentiment vector dimension to the maximum utterance length to achieve the unity of the relative position and absolute position of the sentimental words.

The **sentimental word vector (SWV)** is composed of the word vector and the word-level sentiment vector. Fig. 2 shows the process of the sentimental word encoding. word2vec is used to generate the word vector of "happy". The word sentiment encoder first constructs the word-level sentiment of "happy" and then generates the word-level sentiment vector. SWV of "happy" is the concatenation of the word vector and word-level sentiment vector.

3.1.2 *Utterance Sentiment Encoding*

In order to capture the co-occurrence and dependence between sentimental words in an utterance, the **utterance sentiment vector (USV)** is defined. USV is a vector for describing the utterance-level sentimental information and is calculated by the utterance sentiment encoder on the input utterance.

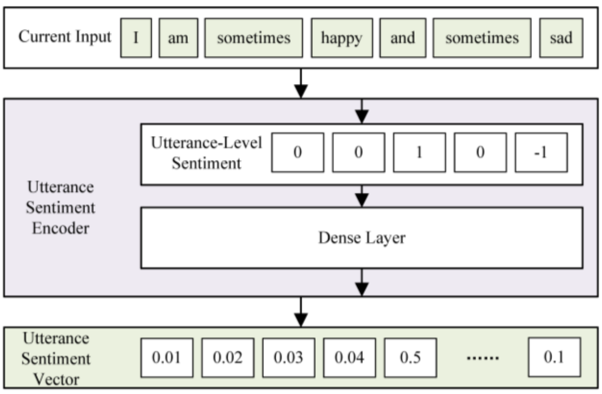

Figure 3. Process of the utterance sentiment encoding.

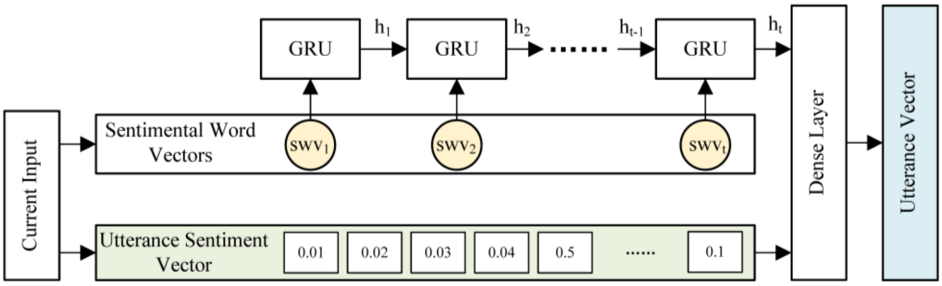

Figure 4. Architecture of the utterance encoder.

Fig. 3 shows the process of the utterance sentiment encoding. The utterance sentiment encoder first generates the utterance-level sentiment of the input utterance and then generates USV.

3.2 Encoding

3.2.1 Utterance Encoding

The utterance of the k-th turn of dialogue $s_k^1 =< w_1, w_2, \cdots, w_i, \cdots, w_l >$, at some time $t(1 \leq t \leq l)$, the hidden state (h_t) of the sentimental word vector (swv_t), and the utterance sentiment vector (usv) of the sentence (s_k^1) generate an utterance vector under the action of the utterance vector function (f_{uv}), expressed as Eq. 1. The final hidden state (h_t) obtained by inputting at time t is generated by the GRU encoder, and is generated by the sentimental word vector (swv_t) at time t and the hidden state (h_{t-1}) at the previous moment, i.e. Eq. 2

$$UV = f_{uv}(usv, h_t) \quad (1)$$

$$h_t = GRU(swv_t, h_{t-1}) \quad (2)$$

The **utterance vector (UV)** is used to describe the semantic and sentimental encoding information of the utterance in the dialogue generation model. Since the sentimental information is embedded in the input of the encoder, the generated utterance vector also contains the sentimental information of the current input sentence.

Fig. 4 shows the architecture of the utterance encoder. The input corresponding SWV sequence obtains a hidden state through GRU and enters the dense layer together with USV of current input and finally generates an utterance vector.

3.2.2 Context Encoding

When in a dialogue generation model, the generated response is not related to the context, sentiment disappearance or sentimental state transition often occurs. Therefore, our context encoder generates a context vector primarily by learning the context of the dialogue. The context vector includes not only a contextual semantic information, but also a contextual sentimental information. The design of our context encoder is inspired by the hierarchical neural network designed by Serban et al. [19].

The **context vector (CV)** is used to describe the semantic and sentimental encoding information of the dialogue context. In the encoding of the k-turn dialogue, the context vector (CV) at the time $T(1 \leq T \leq 2k-1)$ of a certain sentence input is generated by a hierarchical hidden state vector (H_{T-1}) (H_{T-1} is generated by the hierarchical encoder of the 1 to T-1 layer) and the utterance sentiment vector (usv_T) at time T, i.e. Eq. 3.

$$CV = f_{cv}(usv_T, H_{T-1}) \quad (3)$$

where f_{cv} is the context vector generation function. Eq. 4 is the utterance sentiment vector, f_{utt_enc} is the utterance sentiment encoder and s_T is the sentimental word vector sequence generated by the sentence at time T. The k-turn dialogue consists of k sentences of the user, i.e., $d =< s_1, s_2, \cdots, s_i, \cdots, s_k >$. The context hidden state vector (H_{T-1}) of the second layer at time T is generated by the sentence-level hidden state vector (hs_{T-1}) (hs_{T-1} is generated by the sentence at time T-1 calculated by Eq. 5) and the context hidden state vector output (H_{T-2}) at the last time of the second layer, i.e.

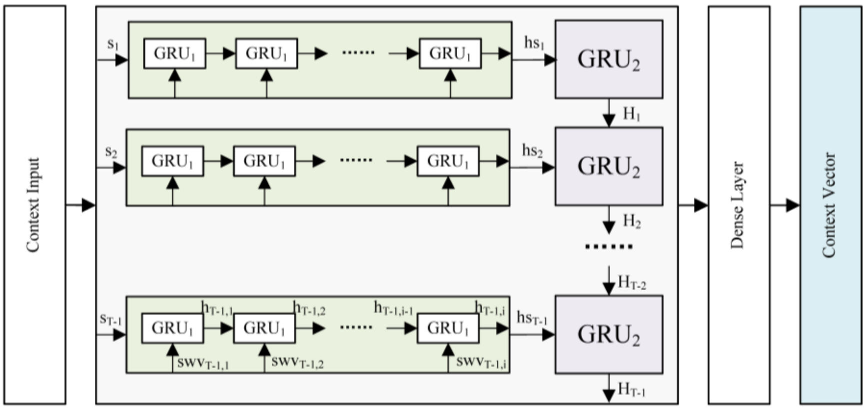

Figure 5. Architecture of the context encoder.

Eq. 6.

$$usv_T = f_{utt\_enc}(s_T) \quad (4)$$

$$hs_{T-1} = \mathrm{h}_{T-1,i} = GRU_1\left(swv_{T-1,i}, h_{T-1,i-1}\right) \quad (5)$$

$$H_{T-1} = GRU_2(\mathrm{h}s_{T-1}, H_{T-2}) \quad (6)$$

Fig. 5 shows the architecture of the context encoder. The dialogue context is the input into the two-layer GRU to obtain a hidden state vector representing the context information. The current input obtains the current USV, and the two enter the dense layer together to generate the context vector.

3.3 Decoding

Our model decoder is a unidirectional GRU neural network. Using the attention mechanism, the concatenation of the utterance vector and the context vector are fed into the decoder to generate a dialogue response.

Bahdanau et al. [20] were the first to propose attention mechanism in the neural machine translation to alleviate the problem of the information loss in the model and to significantly improve the translation quality. The attention mechanism is also widely used in other NLP tasks.

In our model we use the utterance attention mechanism to help the decoder attend important words of an utterance and the context attention mechanism to help the decoder attend important utterances of a context.

4 EXPERIMENTS

4.1 Experimental Setup

To verify the validity of our model, the Douban Chinese dialogue corpus [21] is used. Each row in the corpus represents a set of multiple turns of the dialogue. The HowNet sentiment lexicon, which includes 837 positive and 1254 negative words is used.

In all the experiments, two-layer GRUs are used as an encoder and decoder with 512 hidden size and 256 embedding dimensions. The training is made with the batch size of 32 utterance pairs. Adam with lr = 1e-1 is used for optimization. The dropout rate is set to 0.5.

The used comparative models are the traditional Seq2Seq and the affective neural response generation (ANRG) model. Our model was not compared with ECM, because of its generating specific emotional responses depending on an external vocabulary, which is not our goal.

4.2 Experimental Results and Analysis

Our experimental results are shown in Table 1. The comparison of our model with the Seq2Seq and ANRG models is made in terms of the sentiment, quality and context.

4.2.1 Aspect #1: Sentiment

The metrics needed to evaluate the model sentiment generation ability are: **polar vocabulary (PV)**, **sentimental utterances rate (SUR)** and **polar utterances rate (PUR)**.

PV is a sum of all the polar vocabulary words (including the positive and negative words) in all utterances generated by the model used to measure the ability of the model to generate sentimental words.

SUR is a ratio of the sentimental utterances (the utterances containing the sentimental words) to the total number of the utterances used to measure the distribution of the model on the generation of sentimental words.

Since the above sentimental utterances only access whether they contain sentimental words, they do not tell whether the utterances have sentimental polarity. Therefore, we define a **PUR** to describe the proportion of the utterances with a sentimental polarity in all generated utterances.

In our experiments, 500 dialogue utterances are selected from the test set for evaluation. The sentimental polarity values are calculated using the natural language processing tool package SnowNLP sentiment classifier, and the utterances with the classification probability greater than 0.7 are counted. The experiments show that our model gets the highest sentimental vocabulary, which means that our model has the best performance in sentimental dialogue generation.

Table 1 shows that our model generates the most sentimental utterances on the test set. A sentimental utterance rate is by 60.4% higher than with the traditional Seq2Seq model and by 9.6% higher than with ANRG; the more sentimental utterances, the higher model quality, our goal is to generate responses with more sentimental utterances, so that the multi-turn sentimental dialogue generation model retains the taste of users. The polar utterances rate is by 51.4% higher than with the traditional Seq2Seq model and by 7.8% higher than with ANRG. In daily conversations, the sentimental dialogue does not necessarily have sentimental words, so we use the polar utterances rate

Table 1. Results of different metrics.

Model	Sentiment			Quality				Context	
	PV	SUR (%)	PUR (%)	BLEU-2	BLEU-3	BLEU-4	Average BLEU	Cosine	ConvNet (%)
Seq2Seq	113	20.4	30.2	39.2	29.1	17.2	28.5	0.554	61
ANRG	435	71.2	73.8	39.1	31.4	19.8	30.1	0.552	60
SCS-Seq2Seq	**505**	**80.8**	**81.6**	**40.7**	**32.9**	**21.5**	**31.7**	**0.565**	**64**

evaluation indicators to calculate the proportion of utterances we generate sentimentally. The experimental results show that under the condition of no sentimental words our model still generates sentimental responses.

4.2.2 Aspect #2: Quality

We use BLEU [22] to evaluate the quality of the model generated responses.

BLEU is one of the commonly used automated metrics for machine translation. A higher BLEU value means a higher similarity between the translation sentence and the target translation sentence, indicating a higher quality of the model translation sentence. In these experiments, the BLUE value is used to evaluate the degree of similarity between the generated response of the model and the corpus utterance. In our experiments, we use BLEU-2, BLEU-3, and BLEU-4.

As seen from Table 1, our model is by 3.2 higher than the traditional Seq2Seq model on an average BLEU. Compared with ANRG, the average BLEU is increased by 1.6, which means that our model does not only generate the sentimental but also the syntax utterance. The improvement on the BLEU-4 indicates that our model captures long distance dependencies better than comparative models. The long-distance dependencies can capture the relationship in the generation words.

4.2.3 Aspect #3: Context

We use the text similarity to verify the contextual relevance between the generated response and dialogue context. A higher text similarity indicates a stronger contextual relevance between the generated response and dialogue context. The metrics are the **cosine** similarity and **ConvNet** probability [23].

The **cosine** similarity is one of the commonly used metrics of the context correlation. Its value ranges from -1 to 1. The **ConvNet** probability uses convolutional neural networks (CNN) [24] to combine the differences of CNN with different specifications to measure the similarity of the utterances.

Table 1 shows that the cosine similarity of our model is by 0.011 higher than that of the traditional Seq2Seq model and by 0.013 higher than ANRG. For the ConvNet probability our model is by 3% higher than the traditional Seq2Seq model and by 4% higher than ANRG. This indicates that the responses generated by our model are more relevant to the dialogue context than the comparative models.

4.3 Case Study

To further explain that our model generates the sentimental utterances without simply stacking the sentimental words. "The weather is really good" is token as the input and "Hah, I want to go out" as the output. Fig. 6 shows the probability distribution of the sentimental words generated at different time-steps.

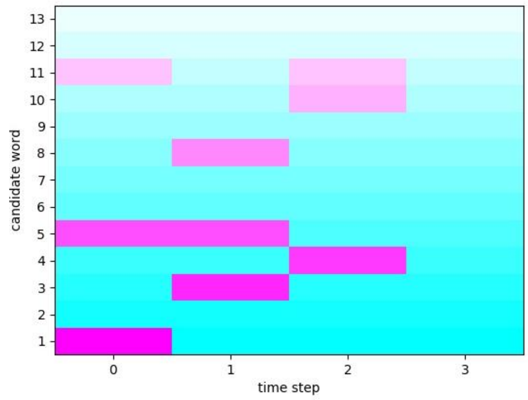

Figure 6. Visual distribution of the sentimental word probabilities.

Fig. 6 shows the sentimental word probabilities of the predicted words. The ordinate represents the candidate words arranged according to the probabilities. The first digit represents the word with the highest probability. In the graph for the first 13 words with the highest probability, the abscissa represents the time-steps and four time-steps are given. The colors denote the probabilities of the candidate words, i.e. the darker the color, the greater the selection probability. The sentimental words are marked in pink and others in blue.

Fig. 6 shows that there is a certain number of the sentimental words in the candidate words generated at each time-step. This means that the probability of generating the sentimental word vocabulary in the generation process is relatively high. At the same time, there are no sentimental words among the first 13 words with the highest generation probability at $t = 3$. This indicates that the model does not simply stack sentimental words in the process of word generation.

Table 2. Examples of responses generated by our model (translated from Chinese).

	Neutral Utterance	Negative Utterance	Positive Utterance
U1	Have you eaten? （你吃饭没？）	This person messed me up. （这人惹到我。）	Take a photo, is it pretty? （晒照片一张，好不好看？）
R1	**Hah** never eaten. （**哈哈**没吃过。）	You are not **happy**. （你不**高兴**。）	The photo is **pretty**. （照片挺**漂亮**。）
U2	Me too. （我也没吃过。）	I am really angry. （我真的生气。）	I'm afraid that something will be seen by others, sorry, hah. （我怕发些东西被人家看到，不好意思，哈哈。）
R2	So **delicious** why don't you **like** it? （这么**好吃**你为什么不**爱**吃？）	Do not be **angry**. （别**生气**。）	Calm down. （你淡定。）
U3	Let's eat together. （那一起吃饭。）	You are so casual, hah. （你好随意，哈哈。）	Every time I look at my own photos, I feel as young as before. （我每次看我自己照片，都觉得和以前一样年轻。）
R3	Come and eat what tastes **good**. （来吃什么口味**好**。）	I have no idea. （我没办法。）	**Just** a little bit. （**差**一点点。）

In Table 2 presenting several examples of responses generated by our model. U is the user utterance, and R is the model-generated response. Sentimental words are marked in bold fonts. As seen, our model generates sentimental responses to both the sentimental and neutral utterances.

5 CONCLUSION AND FUTURE WORK

In the paper we propose a sentimental and context-sensitive Seq2Seq model for dialogue generation. Our experimental results demonstrate that: (1) The sentimental information reinforcement encoder effectively improves the appearance rate of the sentimental words in the dialogue responses and makes the responses contain obvious sentimental polarities; (2) The utterance encoder and context encoder significantly improve the contextual relevance and the quality of the generated responses. Our model generates more sentimental, contextually relevant and higher quality responses.

There are still some shortcomings in our sentimental and context-sensitive Seq2Seq model. One is the lack of sentimentally-oriented control which may lead to a poor sentimental controllability. The focus of our future research will be on controllability and sentimental orientation of the sentimental dialogue generation model.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (No.61762070 and No.61650205) and Natural Science Foundation of Inner Mongolia (No.2019MS06004).

REFERENCES

[1] Turing A M. Computing Machinery and Intelligence[J]. Mind, 1950, 59(236):433-460.

[2] Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine[J]. Communications of the ACM, 1966, 9(1): 36-45.

[3] Inaba M, Takahashi K. Neural utterance ranking model for conversational dialogue systems[C]//Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2016: 393-403.

[4] Shang L, Lu Z, Li H. Neural responding machine for short-text conversation[J]. arXiv preprint arXiv:1503.02364, 2015.

[5] Ritter A, Cherry C, Dolan W B. Data-driven response generation in social media[C]//Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, 2011: 583-593.

[6] Vinyals O, Le Q. A neural conversational model[J]. arXiv preprint arXiv:1506.05869, 2015.

[7] Sordoni A, Galley M, Auli M, et al. A neural network approach to context-sensitive generation of conversational responses[J]. arXiv preprint arXiv:1506.06714, 2015.

[8] Li J, Galley M, Brockett C, et al. A diversity-promoting objective function for neural conversation models[J]. arXiv preprint arXiv:1510.03055, 2015.

[9] Mou L, Song Y, Yan R, et al. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation[J]. arXiv preprint arXiv:1607.00970, 2016.

[10] Li J, Galley M, Brockett C, et al. A persona-based neural conversation model[J]. arXiv preprint arXiv:1603.06155, 2016.

[11] Ghosh S, Chollet M, Laksana E, et al. Affect-lm: A neural language model for customizable affective text generation[J]. arXiv preprint arXiv:1704.06851, 2017.

[12] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems. 2014: 3104-3112.

[13] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.

[14] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model[C]//Eleventh annual conference of the international speech communication association. 2010.

[15] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[16] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.

[17] Zhou H, Huang M, Zhang T, et al. Emotional chatting machine: Emotional conversation generation with internal and external memory[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[18] Asghar N, Poupart P, Hoey J, et al. Affective neural response generation[C]//European Conference on Information Retrieval. Springer, Cham, 2018: 154-166.

[19] Serban I V, Sordoni A, Bengio Y, et al. Hierarchical neural network generative models for movie dialogues[J]. arXiv preprint arXiv:1507.04808, 2015, 7(8).
[20] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
[21] Wu Y, Wu W, Xing C, et al. Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots[J]. arXiv preprint arXiv:1612.01627, 2016.
[22] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002: 311-318.
[23] He H, Gimpel K, Lin J. Multi-perspective sentence similarity modeling with convolutional neural networks[C]//Proceedings of the 2015 conference on empirical methods in natural language processing. 2015: 1576-1586.
[24] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series[J]. The handbook of brain theory and neural networks, 1995, 3361(10): 1995.

Zhiqiang Ma received his B.E. degree in computer application technology from the Hohai University, Nanjing, China, in 1995 and his M.E. degree in computer science and technology from the Beijing Information Science and Technology University, Beijing, China, in 2007. In 1995, he joined the Inner Mongolia University of Technology. Currently he is a Professor at the College of Data Science and Application. He is a reviewer of the Journal of Chinese Information Processing, IEEE ACCESS, and the International Conference of Pioneering Computer Scientists, Engineers and Educators. His research interests include deep learning, speech recognition and natural language processing.

Baoxiang Du received his B.E. degree in computer science and technology from the Qingdao Binhai University, Qingdao, China, in 2018. In the same year, he joined the Inner Mongolia University of Technology. Currently he is a graduate student at the College of Information Engineering. His research interests include deep learning and natural language processing with a special focus on emotional dialogue generation.

Ji Shen is a master student at the College of Data Science and Application, Inner Mongolia University of Technology. He is a CCF member and his research interests include deep learning and natural language processing.

Chunyu Wang received his B.E degree in network engineering from the Taiyuan University, Taiyuan, China, in 2019. In the same year, he joined the Inner Mongolia University of Technology. Currently he is a graduate student at the College of Information Engineering. His research interests include deep learning and natural language processing with a special focus on emotional dialogue generation.

Rui Yang received her B.E. degree in computer science and technology from the Shanxi Agricultural University, Jinzhong, China, in 2017. In the same year, she joined the Inner Mongolia University of Technology. Currently she is a graduate student at the College of Information Engineering. Her research interests include deep learning and natural language processing with a special focus on emotional dialogue generation.