Text Generation Based on Empathetic Dialogues between Nurses and Patients

Wenxu Shi, Zuyue Shang, Shengli Bao and Guoyong Li
Chengdu Institute of Computer Application Chinese Academy of Sciences
University of Chinese Academy of Sciences Chengdu, China
[a]shiwenxu1995@126.com, [b]764304546@qq.com,
[c]baohigh@casit.com.cn, [d]liguoyong@casit.com.cn

Abstract. The communication between nurses and patients plays an important role in medical work. Perception and expression of emotions are key factors in the nurse-patient dialogue system. Using empathy between nurses and patients can make their communication more effective. With the development of artificial intelligence, an important research direction of computer-aided diagnosis is to enable computers to perceive the emotions of nurses and patients, and to make situational responses to specific emotions. This paper, we present a virtual machine model based on deep learning, which is used to simulate the emotional expression of empathy dialogue between nurses and patients. Firstly, the emotional expression between nurses and patients is simulated by embedding the text dialogue of emotion category, then the attention mechanism is used to capture the changes of emotion in the text dialogue, and finally, the words with emotion are used to express the emotion explicitly. The results of the experiment indicate that the proposed model can not only generate the content in accordance with the dialogue scenario, but also generate emotions.

Keywords: Emotion analysis, empathy, man-machine dialog, emotional perception.

1. Introduction

The ability to perceive and express emotions is one of the most important cognitive behaviors of human beings[1,2]. With the development of society, the demand for medical care has expanded from simple disease treatment to medical technology, medical environment, service attitude, and even higher spiritual needs. Communication between nurses and patients is a concrete manifestation of human-computer communication ability in medical work. It is an important research direction of computer aided diagnosis to let the computer perceive the emotions of the nurses, understand the emotions of the nurses and make the corresponding emotions. Studies have shown that the level of empathy directly affects the quality of care, and empathy is a key factor for effective communication and understanding between nurses and patients [3, 4]. The word 'empathy' can be translated into many ways In psychology, not sympathy, but reasonable, understanding, which is the ability to put yourself in their shoes and try to understand why they think the way they do[5,6].

We design a sequence-sequence generation model with emotional perception and expression mechanism for nurse-patient empathy. The emotional expression between nurses and patients is simulated by embedding the text dialogue of emotion category, then the attention mechanism is used to capture the changes of emotion in the text dialogue, and finally, the words with emotion are used to express the emotion explicitly. The proposed model can not only generate content in accordance with the conversation situation, but also generate emotions, which achieves the desired experimental objectives.

In summary, this paper makes the following contributions:

An emotional interaction model for nurse-patient empathy was proposed.

The emotion category is embedded in the dialogue and the perception and expression of emotion are realized.

The proposed model can not only generate the content in accordance with the dialogue scenario, but also generate emotions.

2. Related Work

Perception and expression of emotions has always been the focus of human-computer interaction research. Analyzing the emotion of dialogue between nurses and patients and making appropriate response to it can improve the satisfaction of patients. In 1966, Joseph Weizenbaum, a professor at the Massachusetts Institute of Technology (MIT), developed a chat robot called ELIZA, which was used to simulate a psychiatrist in clinical treatment. This is the first time that a chat robot has been used in medical treatment career [7].In 1995, a chat robot named A.L.I.C.E (Alicebot) was developed and won the European Artificial Intelligence Award several times [8]. In 2005, Prendinger [9] embedded emotional words into conversation, and concluded that embedding emotions in conversation system can significantly reduce user pressure and enhance user satisfaction. Skowron [10] proposed a conversation system named "emotional audience" in 2010, which aims at detecting and adapting users'emotional state, and responding meaningfully to users' words in terms of content and emotion. In 2017, Cagan et al. [11] combined grammatical information to generate comments for documents using emotions and topics. In 2017, Ghosh et al. [12] proposed an emotional language model to generate text based on contextual and emotional categories. On the basis of predecessors, we proposed an empathetic text generation for the nurse-patient dialogue.

3. Data Preparation

It is a challenge to solve the empathy factor in large-scale session generation. Emotional tagging for dialogue is a fairly subjective work and it is difficult to obtain high quality emotional tagging data in large-scale corpus. Since there is no available data for empathy dialogue nowadays, we trained several classifiers on the NLPCC2017 dataset, and then selected the classifier with the best classification accuracy to automatically annotate the conversations gathered from the network.

3.1 MLP

Multilayer Perceptron (MLP) is also called Artificial Neural Network. In addition to the input and output layers, there are many hidden layers in the middle. The main purpose of the hidden layer is to enhance the expression ability, the more layers of the hidden layer, the higher the natural expression ability. The simplest MLP contains only one hidden layer, namely three layers. The simplest structure of multi-layer perceptron is shown in Fig. 1

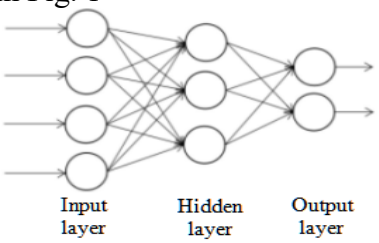

Fig. 1 The simplest structure of multi-layer perceptron

The process can be formulated as follows:

$$f(x) = G(b^{(2)} + W^{(2)}(s(b^{(1)} + W^{(1)}x))) \tag{1}$$

Where x is the input element of the input layer, $W^{(i)}$ is the weight parameter of each layer, $b^{(i)}$ is the parameter of each layer offset, and the function $s(\cdot)$ usually adopts the following two functions:

$$sigmoid(a) = \frac{1}{1+e^{-a}} \tag{2}$$

$$\tanh(a) = \frac{e^{a} - e^{-a}}{e^{a} + e^{-a}} \tag{3}$$

The function $G(\cdot)$ can be formulated as follows:

$$soft\max(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{i=n} e^{z_i}} \tag{4}$$

3.2 LSTM

Long Short-Term Memory (LSTM) not only solves the problem of long-distances dependencies, but also solves the problem of gradient vanishing and explosion in the traditional RNN model [13]. The LSTM network structure is shown in Fig. 2

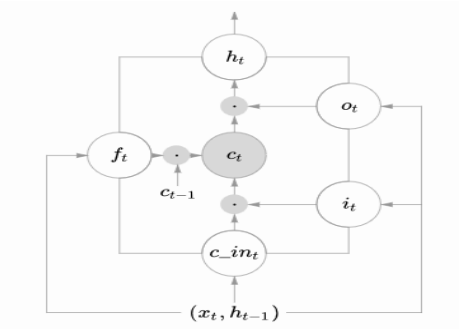

Fig. 2 The structure of LSTM network

The LSTM model consists of a series of repetitive timing modules, at time t, the input gate, the forget gate, and the output gate, denoted as f_t, i_t and o_t respectively. The process can be formulated as follows:

$$f_t = \partial(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

$$i_t = \partial(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$o_t = \partial(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$c_t = f_t \Theta c_{t-1} + i_t \Theta \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{8}$$

$$h_t = o_t \Theta \tanh(c_t) \tag{9}$$

Where W_f、W_i、W_o, and W_c are the weight matrices that map the input of the hidden layer to the state of the three gates and the input unit. The b_f、b_i、b_o and b_c are used to represent four bias vectors, respectively. The $\partial(\cdot)$ is the activation function, which usually uses the sigmoid function to perform the operation. The $\tanh(\cdot)$ is the hyperbolic tangent function, and the Θ is used to represent the dot product operation.

3.3 Bi-directional LSTM

One shortcoming of conventional LSTM is that they can only take advantage of the previous context, but bidirectional LSTM (Bi_LSTM) processes bidirectional data through two independent hidden layers, and then fed forwards the two hidden layers to the same output layer, thus overcoming the shortcomings of traditional LSTM[14]. Bi_LSTM considers the context of the text at the same time. Fig. 3 illustrates the structure of an expanded Bi_LSTM layer, which consists of a forward LSTM layer and a backward LSTM layer.

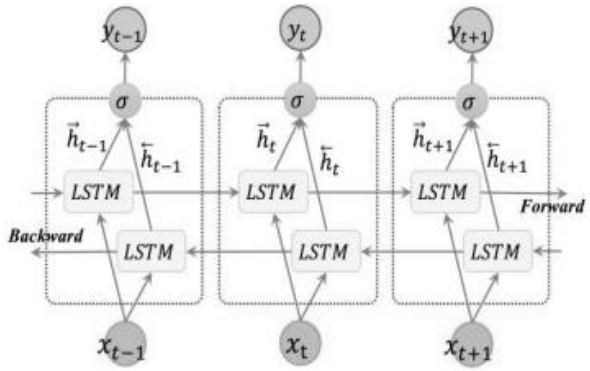

Fig. 3 Unfolded architecture of bidirectional LSTM

The hidden state H_t of Bi_LSTM at time t, including forward $\vec{h}_t$, and backward $\overleftarrow{h}_t$:

$$\vec{h}_t = \overrightarrow{LSTM}(h_{t-1}, w_t, c_{t-1}), t \in [1,T] \tag{10}$$

$$\overleftarrow{h}_t = \overleftarrow{LSTM}(h_{t+1}, w_t, c_{t+1}), t \in [T,1] \tag{11}$$

$$H_t = [\vec{h}_t, \overleftarrow{h}_t] \tag{12}$$

3.4 Bi_LSTM+Attention

The Attention mechanism simulates the characteristics of human brain that can capture important information, so that the model can capture the most important parts of the sentence under the consideration of different aspects. The Attention mechanism has been applied in many fields, such as Xu[15] introduced an attention based model, which can automatically learn to describe the content of an image. Mnih [16] proposed an attention model based on neural network, which can select a series of regions or locations autonomously, and only process the selected regions at high resolution, so as to extract information from images or videos. According to the characteristics that attention mechanism can focus on different parts of sentences, Wang [17] proposed an emotional classification model based on attention-based long and short memory networks. Fig. 4 is the model structure of attention mechanism.

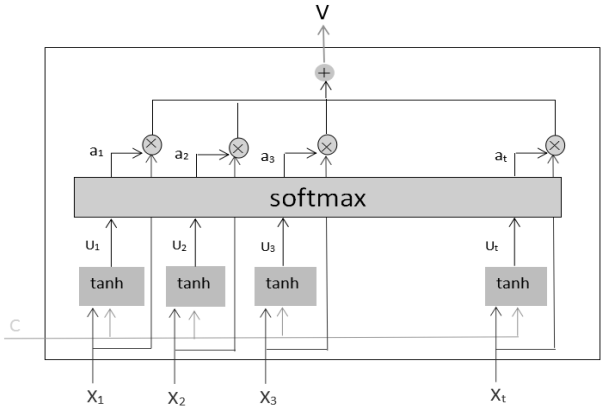

Fig. 4 The structure of attention mechanism

$$u_t = \tanh(w_w x_t + b_w) \tag{13}$$

$$a_t = soft\max(u_t^T, u_w) \tag{14}$$

$$v = \sum_t a_t x_t \tag{15}$$

In the above formula, u_t is the hidden unit of x_t, u_w is the context vector, a_t is the attention vector, v is the output vector through the attention mechanism. u_w is a random initialization and continuous learning in the training process.

3.5 Classification

According to the four classifiers mentioned above, we trained in NLPCC2017 dataset, Which is manually annotated with 6 emotion categories: Angry, Disgust, Happy, Like, Sad, and Other. We then

partitioned the NLPCC2017dataset into training and validation, with the ratio of 8:2, the statistics of it are shown in Table 1.

Table 1 Statistics of the NLPCC2017 dataset

	Others	**Like**	**Sad**	**Disgust**	**Angry**	**Happy**
Train	213567	183608	124055	154158	88124	128487
Test	53354	46031	31183	38442	21884	32105

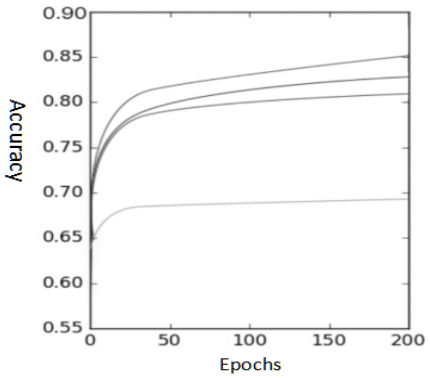

Fig. 5 Classification Accuracy of NLPCC2017 dataSet

Table 2 Classification accuracy on the NLPCC2017

Method	**Accuracy**
MLP	68.55%
LSTM	80.86%
Bi_LSTM	82.90%
Bi_LSTM+Attention	85.68%

The results in Fig. 5 and Table 2 show the classification results of all the neural classifiers, and the Bi_LSTM+Attention classifier achieves the best accuracy of 85.68%. Then, we applied the best classifier, Bi_LSTM+Attention, to annotate six emotion categories for Datasets collected from the web. The statistics on the distribution of data on different emotions in training and validation sets are shown in Table 3.

Table 3 Statistics of the Dataset

	Category	Train	Validation
Conversation	Others	198069	16362
	Like	200001	15247
	Sad	181252	7727
	Disgust	200001	10727
	Angry	139883	4134
	Happy	200001	6788

4. Emotional Chatting Machine

4.1 Encoder-Decoder Framework

The Encoder-Decoder framework can be understood as a generic processing model that generates another sentence from one sentence. For the sentence pair <X, Y>, our goal is to give the input sentence X, expecting to generate the target sentence Y through the Encoder-Decoder framework. X and Y are each composed of a sequence of words.

$$X =< X_1, X_2 \cdots X_m >$$
$$Y =< Y_1, Y_2 \cdots Y_n >$$

Our model is based on an encoder-decoder framework, which combines sequence-to-sequence (seq2seq) with attention structure. The structure is shown in Fig. 6.

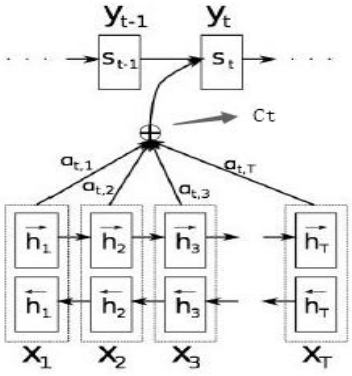

Fig. 6 Attention mechanism in Encoder-Decoder framework

The model with Attention mechanism in Encoder-Decoder framework can be summarized as follows：

$$p(y_t \mid y_1, y_2, ..., y_{t-1}, x) = g(y_{t-1}, s_t, c_t) \tag{16}$$

$$s_t = f(s_{t-1}, y_{t-1}, c_t) \tag{17}$$

$$c_t = \sum_{j=1}^{T_x} a_{tj} h_j \tag{18}$$

$$a_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^{T_x} \exp(e_{tk})} \tag{19}$$

$$e_{tj} = score(s_{t-1}, h_j) = v^T \tanh(W_{s_{t-1}} + U_{h_j}) \tag{20}$$

Where s_t is the state of the hidden layer of step t in the output network, h_j is the state of the hidden layer of step j in the input network, a_{tj} is the weight. In the t th step of the output network, the sum of a_{tj} is 1. e_{tj} is to measure the correlation between s_{t-1} and h_j. If the higher the correlation between s_{t-1} and h_j, the more attention should be focused on the elements near the j position in the input and give them a higher weight. In different papers, the calculation method of e_{tj} is different. For details, please refer to the literature [18, 19].

4.2 General Framework

First, we get tens of thousands of conversational data from the network, including a post and a response, this dataset is good enough to train the models in practice. Then we use the previously trained sentiment classifier (Bi_LSTM+Attention classifier, accuracy 85.68%) to automatically mark the emotional categories of posts and responses, which will be labeled as one of the categories of "happy, like, sad, disgusted, angry, and other". These marked conversation data will be sent to the emotional expression module to get a response to the corresponding emotion based on the sentiment of the posting. Finally, we will judge the sentiment type of the response sentence obtained to judge the empathy accuracy of the generated statement. The overall framework is shown in Fig. 7.

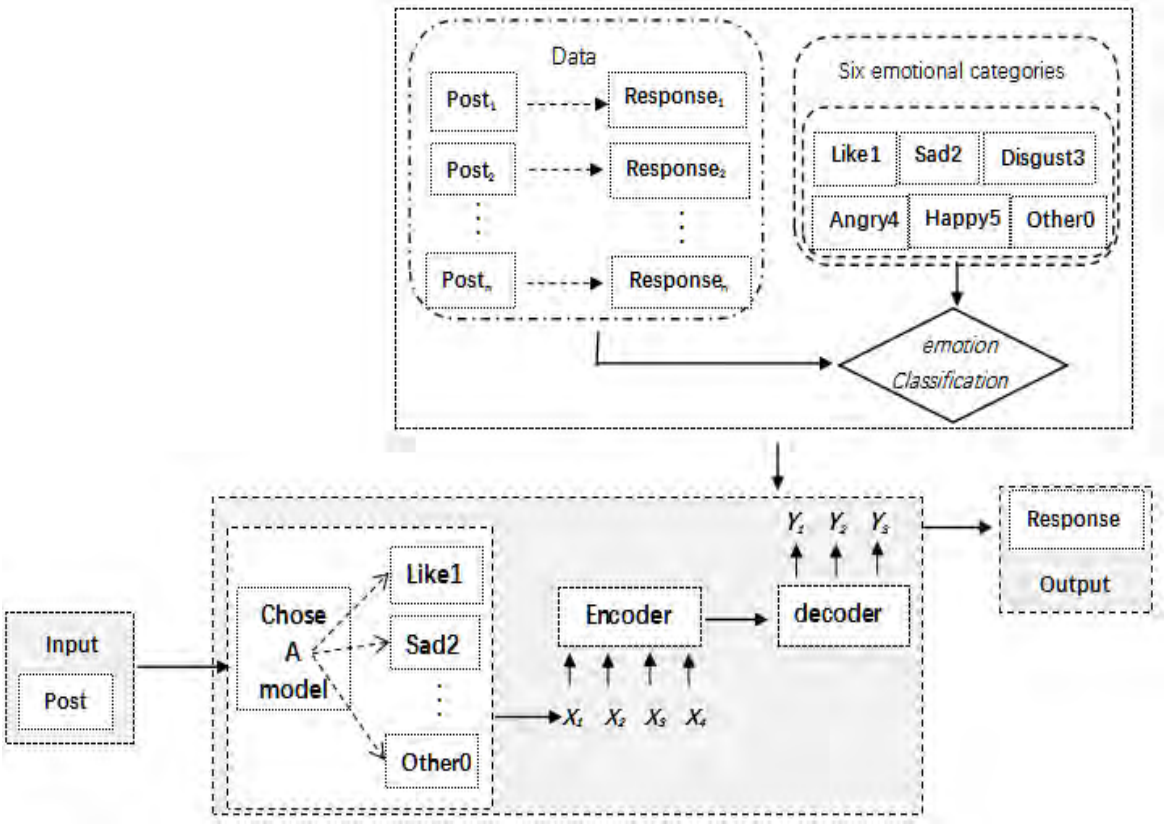

Fig. 7 Overview of general framework

4.3 Experimental Result

Table 4 Accuracy of empathy for different emotional categories

Category	Test	Acc
Others	1534	0.962
Like	1365	0.966
Sad	1430	0.994
Disgust	1256	0.934
Angry	1196	0.904
Happy	1432	0.990

In order to judge whether the sentences generated by the model have the characteristics of empathy, for each kind of emotion, we collected several questions to test the model. If the emotional category of the response is the same as the post, we believe that the response is empathic. The statistics of the empathy accuracy rate of different categories of sentiment data are shown in Table 4.

We present some examples in Table 5. The words marked red are emotional words corresponding to a given emotional category. It can be seen that for a given post, the model can not only generate content in accordance with the conversation situation, but also generate emotions.

Table 5 Sample responses generated by our model

category	Post(translated)	response(translated)
like	I looked at it, and I thought it was good.	I think it's pretty good.
	I expect you to find out how much I love you.	I love you too.
	Be my girlfriend, I will make you happy.	I'll make you happy, too.
sad	I have a sore throat and swollen gums.	You're so pathetic.
	It's been two hours. why hasn't anything shown up yet? I'm so tired.	Oh, I don't want that to happen.
	I was in a bad state of mind this morning because I didn't sleep well last night!	I was in a bad state of mind this morning, too.
disgust	When I think of the food on the plane today, I want to vomit.	It's so disgusting!
	You already know, why not say it earlier? it's a loud hindsight.	It's my problem. don't take it seriously. don't say that about me.
	Give me some reason to pay attention to Chinese football! Hurt me again and again.	Stop paying attention to Chinese football. it's disgusting to think about it.
angry	I've been waiting for you to eat for three hours.	Isn't it right to wait for us to eat?
	I will never look for you again.	What did you say? When are you coming!
	It is your money that irritated me!	You are an idiot! I called you but you didn't pick it up!
happy	Happy birthday, I love you.	Aha!
	Sister should formally wish you a happy holiday! Get what you want!	Ha, thank you for your blessing! I miss you.
	You are so beautiful!	Haha, thank you.
	The good news of their marriage made me want to get married, too.	Ha, you come here and we'll get married.
others	Where's the boss?	He's at the Beijing International Convention and Exhibition Center.
	What time does this start?	Let's make it Wednesday for the time being.
	What time does the teacher come back?	The teacher will be back on Sunday.

5. Conclusion and Future Work

This paper explores an empathic textual dialogue of nurse-patient relationships, which shows that computers can express initial emotions. Through the training of a large number of conversational corpus, the computer can respond to different emotions, which can be used in more complex medical tasks such as emotional care, psychological counseling and so on. In perception and expression of emotions，robots should be able to recognize, understand and express human emotions from multi-modal information, such as text, voice and expression. This requires the comprehensive

application of natural language understanding, speech recognition and synthesis, visual information processing, robot planning and execution and other technologies.

In the future work, we will explore how to predict the emotional content that the computer is most suitable for in response to the user's input and the current situation.

Acknowledgments

This paper was supported by the Key Research and Development Projects of Science and Technology department of Sichuan province (2018SZ0040), the new generation of AI major projects of Science and Technology department of Sichuan province (2018GZDZX0036).

References

[1]. Salovey P, Mayer J D. Emotional intelligence [J]. Imagination Cognition and Personality. 1990, 9(3):185– 211.

[2]. Picard R W, Picard R. Affective computing [M]. Cambridge: MIT press, 1997.

[3]. Prendinger, H., and Ishizuka, The empathic companion: A character-based interface that addresses users ’affective states. [M] Applied Artifi cial Intelligence 2005,19(3-4):267 – 285.

[4]. Dong Dandan. Application of incentive mechanism in orthopedic care management [J]. Chinese Journal of Aesthetic Medicine (Comprehensive Edition), 2011, 3(6):485.

[5]. Guo Nianfeng. National Vocational Qualification Training Course. Psychological Counselor. Level 3 [M]. Beijing: Ethnic Publishing House, 2005:60.

[6]. Wang Changhong. Clinical Psychotherapy [M]. Beijing: People's Military Medical Publishing House, 2001:55-82.

[7]. 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.

[8]. Wallace R. The Anatomy ofALICE[EB/OL].http://www.alicebot.org/anatomy.html.2006-12-10.

[9]. Prendinger, H. Mori,J. Ishizuka, M. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. International journal of human-computer studies 2005,62(2):231 – 245.

[10]. Skowron, M. Affect listeners: Acquisition of affective states by means of conversational systems. In Development of Multimodal Interfaces: Active Listening and Synchrony. Springer. 2010. 169 – 181.

[11]. Cagan, T. Frank, S. L. Tsarfaty, R. Data-driven broad-coverage grammars for opinionated natural language generation (onlg). In ACL,2017 (1), 1331 – 1341.

[12]. Ghosh S, Chollet M, Laksana E, Morency, L.; and Scherer, S. Affect-lm: A neural language model for customizable affective text generation. In ACL, 2017, 634 – 642.

[13]. Yousfi S,Berrani S, Garcia C. Contribution of recurrent connectionist language models in improving LSTM-based Arabic text recognition in videos.[J].Pattern Recognition ,2017,41(5):245-254.

[14]. Graves, Alex，Mohamed, Abdel-rahman，Hinton, Geoffrey. Speech recognition with deep recurrent neural networks. Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.2013:26-31.

[15]. Kelvin Xu, Jimmy Ba, Ryan Kiros, et al.Show, attend and tell: Neural image caption generation with visual attention[C]. Proceedings of the 32nd International Conference on Machine Learning (ICML), 2015: 2048-2057.

[16]. Mnih V, Heess N, Graves A. Recurrent Models of Visual Attention[C].Advances in Neural Information Processing Systems 27(NIPS), 2014:2204-2212.

[17]. Wang Y, Huang M, Zhu X, et al. Attention based LSTM for Aspect-level Sentiment Classification [J].Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing (EMNL), 2016 : 606-615.

[18]. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In ICLR. 2015.

[19]. Sutskever I, Vinyals O, Le. Q V. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems,2014. 3104 – 3112.