Knowledge-Based Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Topic-enhanced emotional conversation generation with attention mechanism

Yehong Peng, Yizhen Fang, Zhiwen Xie, Guangyou Zhou *

School of Computer, Central China Normal University, Wuhan 430079, China

HIGHLIGHTS

- We present a topic-enhanced neural emotion conversation generation model (TE-ECG) with attention mechanism.
- The topic words are obtained from a pre-trained Twitter LDA model to ensure the generated response is related to the post.
- A novel dynamic emotional attention mechanism is proposed to capture the emotional context and topic information.
- The TE-ECG model can generate responses at both the emotion- and content-related levels.

ARTICLE INFO

Article history:
Received 23 December 2017
Received in revised form 3 September 2018
Accepted 6 September 2018
Available online xxxx

Keywords:
Emotional conversation
Topic model
Sequence-to-sequence
Attention mechanism

ABSTRACT

Emotional conversation generation has elicited a wide interest in both academia and industry. However, existing emotional neural conversation systems tend to ignore the necessity to combine topic and emotion in generating responses, possibly leading to a decline in the quality of responses. This paper proposes a topic-enhanced emotional conversation generation model that incorporates emotional factors and topic information into the conversation system, by using two mechanisms. First, we use a Twitter latent Dirichlet allocation (LDA) model to obtain topic words of the input sequences as extra prior information, ensuring the consistency of content between posts and responses for emotional conversation generation. Second, the system uses a dynamic emotional attention mechanism to adaptively acquire content-related and affective information of the input texts and extra topics. The advantage of this study lies in the fact that the presented model can generate abundant emotional responses, with the contents being related and diverse. To demonstrate the effectiveness of our method, we conduct extensive experiments on large-scale Weibo post–response pairs. Experimental results show that our method achieves good performance, even outperforming some existing models.

1. Introduction

Constructing a short-text emotional conversation system is one of the most challenging and meaningful tasks in natural language processing (NLP) [1,2] and information retrieval (IR) [3,4]. This task aims to generate an utterance that is coherent with a given post and a specified emotion category [1]. Recently, with the popularity of deep neural networks in various NLP and IR applications, conversation generation has become an active research topic in both academia and industry. Emotional intelligence, which is a crucial aspect of human intelligence, comprises both the ability to recognize a person's emotional state and the capability to handle it [5]. In human–computer interaction, the ability of computers to perceive human emotions and accurately handle them can enrich interactions and increase user satisfaction [6]. Hence, conversation systems capable of emotional intelligence are necessary.

Studies conducted on conversation task in past decades can be categorized into the following three groups: rule-based [7,8], retrieval-based [3,4], and statistical machine translation (SMT)-based [9] methods. In the first method, conversational models can directly benefit from rules that depend on manual efforts. However, these models are often effective only in very limited domains. In the second type of method, IR methods are leveraged to select a suitable reply from an existing repository [3,4]. Meanwhile, SMT-based methods regard response generation as a single language translation task [9]. However, we have observed that the last two types are either limited to pre-constructed databases or simply semantically equivalent to a post.

With the growing popularity of social media (e.g., WeChat, Weibo, and Twitter), considerable conversational data have been accumulated on the Web, increasing new opportunities to address a conversation generation problem on large-scale social media. Most existing studies in the literature [2,9–11] have used data-driven methods to improve the content quality of conversational

* Corresponding author.
E-mail addresses: yhpeng@mails.ccnu.edu.cn (Y. Peng), yzfang@mails.ccnu.edu.cn (Y. Fang), xiezhiwen@mails.ccnu.edu.cn (Z. Xie), gyzhou@mail.ccnu.edu.cn (G. Zhou).

https://doi.org/10.1016/j.knosys.2018.09.006
0950-7051/

generation. Recently, several preliminary studies have attempted to generate emotional responses on social media [1,12]. Zhou et al. [1] proposed an emotional chatting machine (ECM) to generate emotional responses based on emotion category embedding, with internal and external emotion mechanisms. Asghar et al. [12] proposed to generate affective conversational texts in three novel manners. However, these methods of emotional conversation generation only focused on emotional factors while ignoring content relevance, possibly resulting in a decline in the quality and diversity of a response. For example, in Table 1, given the post "What's wrong with this? Not calm at all!", the ECM generates a response " I am forced!" Although the ECM can provide a good reply to the post, the content of the response is clearly not relevant. To improve the quality and relevance of emotional responses, we construct a topic-enhanced emotional conversation generation (TE-ECG) model that can simultaneously maintain the relevance of contents and high-quality responses and specify emotion for the reply. As shown in the previous example, TE-ECG can consider emotion-related topic information and generate a further relevant and informative response with the specified emotion "Do you want to say that I am angry?".

TE-ECG is based on an encoder–decoder framework with a dynamic emotional attention mechanism. Our model can generate content-suitable and emotion-related utterances to effectively respond to users' utterances. The generated responses are close to those of daily human conversation. In our framework, we consider two mechanisms – topic module and dynamic emotional attention – to ensure content consistency and emotion suitability of the responses. The topic module is used to acquire a related topic that ensures coherence of conversational interaction and high-quality responses. Meanwhile, the dynamic emotional attention mechanism aims to generate more diverse responses and provide a system with emotional intelligence. To validate the effectiveness of our model, we conduct experiments on large-scale Weibo post–response pairs and compare the presented model with several baselines by using both automatic and human evaluation. The experimental results show that our method can generate content-related and emotional responses and achieve good performance on evaluation metrics, even outperforming some existing models.

The novelty of this study lies in the fact that our method can simultaneously integrate emotional and topic factors in conversation generation. Generally, the main contributions of the study are as follows:

- We present a topic-enhanced emotional conversation generation (TE-ECG) model. The topic module integrates topic factors into responses as prior knowledge to ensure content consistency. Meanwhile, dynamic emotional attention is used to generate replies in both content- and affect-related levels.
- Our system can generate emotional and diverse responses, outperforming those of popular sequence-to-sequence (seq2seq) conversation generation models.

The rest of this paper is organized as follows. In Section 2, related work of the conversation system and emotional dialogue generation is presented. In Section 3, the details of the seq2seq structure and our TE-ECG model are discussed. In Section 4, the dataset, implementation details, and experimental results are provided. Finally, in Section 5, conclusions of this study are elaborated and the future work is revealed.

2. Related work

In this section, we first provide an overview of conversational systems and then describe emotional conversation systems and the distinction of the proposed model.

2.1. Conversation systems

Previous studies on conversation were based on rules or templates [7,8]. Rule-based conversational models require no or few data for training and directly benefit from the rules that depend on substantial manual efforts. However, this type of method is often designed for specific domains only. With the popularity of social media, considerable conversational data are available. Thus, researchers have begun to adopt pure data-driven retrieval-based [3,4,13] or SMT-based [9] methods to address conversational problems. Ji et al. [4] employed the IR techniques to perform conversation tasks, while Ritter et al. [9] proposed a phrase-based SMT approach to generate responses. However, these studies could not adapt well to various contexts.

We follow a new line of investigation, that is, using neural networks to map sequences to sequences [2,10,14–16]. The success of seq2seq generation in machine translation (MT) has inspired researches to attempt to apply these neural models to conversation generation. Serban et al. [14] proposed a hierarchical neural network-based generative architecture, and Sordoni et al. [17] integrated contextual information into a seq2seq model. Other researchers have also integrated deep learning techniques with existing methods. Yan et al. [18] proposed to learn response ranking and next utterance suggestion jointly based on a dual long short-term memory (LSTM) chain model. Ghazvininejad et al. [13] presented a knowledge-grounded neural conversation model that retrieved contextually relevant facts to produce further contentful responses.

Dialogue systems tend to generate several general and meaningless responses, such as "Me too,". Hence, considerable research has been dedicated to improving the content quality of conversation. Mou et al. [10] introduced keywords to replies. Xing et al. [11] combined topic information with a seq2seq framework, while Li et al. [19] presented persona-based models to handle speaker consistency. Other researchers applied deep reinforcement learning [19], adversarial training [20], and beam search [21–23] to generate further diversified and significant responses. Although these studies can generate meaningful and good-quality responses, a chat bot still cannot naturally communicate with users due to a lack of emotional intelligence.

2.2. Emotional dialogue generation

To construct dialogue systems that can chat with users more naturally, several studies have considered emotional factors, yet these systems are mostly limited to small-scale data, only focusing on incorporating emotion into rule-based conversation systems [5,6,24,25]. Recently, researchers have attempted to generate emotional responses based on the encoder–decoder framework in large-scale conversation generation [1,12,26,27]. [12] proposed an affective response generation model to generate sentences conditioned on emotional word embeddings, affective objective functions, and diverse beam search. Zhou et al. [1] proposed the ECM to incorporate emotional factors into conversation generation, utilizing emotion category embeddings, internal emotion states, and external emotion vocabulary. Zhang et al. [27] used multi-task seq2seq learning to construct an emotional conversation system. Besides neural network-based models, several studies exist on emotional conversation that attempt to combine different types of methods [28]. For example, Zhuang et al. [28] presented an ensemble of retrieval- and generation-based open-domain dialogue systems.

Prior studies on emotional dialogue generation focused mostly on emotional factors while ignoring content quality and relevance, possibly resulting in content quality decline. To address this issue, we present a framework to enrich the content and specify emotion to responses simultaneously by incorporating topic

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.

Table 1
Conversations with/without considering the relevance of contents. ECM is one of our baselines, as mentioned in the following section. The informative words from TE-ECG are coloured.

Post	Model	Response	Emotion
好温暖的一幕 A warm scene	TE-ECG	是啊，很温暖 Yes, very warm	Like
	ECM	是的，很有爱的感觉 Yes, a feeling of love.	Like
八嘎你让老娘成功抓狂 Idiot, you make me crazy successfully	TE-ECG	我已经奔溃了！我现在只想睡觉。 I have already collapsed, I just want to sleep now.	Sad
	ECM	我已经奔溃了！！ I have already collapsed!!	Sad
似乎很伤感啊。 It seems very sad.	TE-ECG	不伤感，只是觉得自己太无聊了 Not feel sad, just feel too boring!	Disgust
	ECM	我觉得这个社会太现实了 I think this society is too realistic.	Disgust
这是怎么的了呢，这么不淡定！ What's wrong with this? Not calm at all!	TE-ECG	你是想说我怒了么？ Do you want to say that I am angry?	Angry
	ECM	我是被逼的！ I am forced!	Angry
我是想知道法国考的驾照中国能用么 I want to know if the French driver's license can be used in China.	TE-ECG	哈哈，我看行，我也要考驾照 Haha, I think it's ok, I also want to get my driver' s license.	Happy
	ECM	哈哈，你可以试试. Haha, you can have a try.	Happy

and emotional factors into a conversation generation system. The presented model is inspired by previous studies [1,11], but is unique in that we consider both emotional factors and content relevance in generating responses. Several other studies related to emotion based on a topic model exist [29–31], among which Rao et al. [30,31] aimed to discover topics that convey strong social emotions through emotion topic models. Meanwhile, our study has an essential difference with them in terms of the task and model used. Specifically, our study involves two distinct features. First, we use topic words obtained using a Twitter LDA model as extra prior information to ensure that content consistency is not neglected for emotional conversation generation (ECG). Second, we inject the dynamic emotion information into the text and topic attention respectively, incorporating the content with emotional factors in the decoding process to generate a topic-related and emotional response.

3. Methods

This task is designed to generate a response $Y = (y_1, y_2, \ldots, y_N)$ that is coherent with a given post $X = (x_1, x_2, \ldots, x_T)$ and user-specified emotion category e. The emotion categories include {*Angry, Disgust, Happiness, Like, Sadness, Other*} [32], which are obtained from a bidirectional long short-term memory (BiLSTM) model [1,33]. The dataset includes emotion labels for each post and response. Additional details will be provided in Section 4.

To address this problem, we construct a TE-ECG model. As shown in Fig. 1, our system incorporates emotion and topic modules into a seq2seq framework, and generates emotional expressions by using two mechanisms: topic module and dynamic emotional attention. First, topic words of an input sequence are obtained from a Twitter LDA model as an additional input. Then, dynamic emotional attention, which includes emotion-related text and topic attention, is utilized to adaptively acquire important affective information of the input texts and extra topics.

3.1. Encoder-decoder with attention mechanism

In recent years, many studies have been conducted to generate sentences or responses based on recurrent neural networks (RNNs) [10,14,15,34]. These systems adopted a standard encoder–decoder framework based on seq2seq learning. The encoder initially accepts words in the input sequence sequentially and then encodes the input as a context vector. The output sequence is finally generated from the decoder by using the context vector.

Here, we adopt an optimized encoder–decoder framework, as proposed in [34]. The encoder is based on BiLSTM, including forward and backward LSTMs. The forward LSTM encodes the input sequence $(x_1, x_2, \ldots, x_T)$ in forward direction as a sequence of forward hidden states, $\overrightarrow{h} = (\overrightarrow{h}_1, \overrightarrow{h}_2, \ldots, \overrightarrow{h}_T)$, while the backward LSTM reads the input sequence in reverse order $(x_T, x_{T-1}, \ldots, x_1)$ and calculates a sequence of backward hidden representations, $\overleftarrow{h} = (\overleftarrow{h}_1, \overleftarrow{h}_2, \ldots, \overleftarrow{h}_T)$. The final hidden state of each word x_j, denoted as h_j, is obtained by concatenating the forward and backward hidden states, that is, $h_j = [\overrightarrow{h}_j; \overleftarrow{h}_j]$.

The hidden state at time j can be defined as

$$h_j = \text{BiLSTM}(h_{j-1}, x_j). \tag{1}$$

Then, with the attention mechanism introduced in [34], a context vector c_i is calculated as a weighted sum of the hidden states as follows:

$$c_i = \sum_{j=1}^{T} \alpha_{ij} h_j, \tag{2}$$

where the weight α_{ij} of each h_j is obtained using a softmax function described as

$$\alpha_{ij} = \frac{\exp(\eta(s_{i-1}, h_j))}{\sum_{k=1}^{T} \exp(\eta(s_{i-1}, h_k))}, \tag{3}$$

where η is a multilayer perceptron.

For the decoder part, we adopt a unidirectional LSTM. The context vector c_i, the previously decoded word embedding y_{i-1}, and the hidden state of previous time s_{i-1} are simultaneously fed into the decoder to calculate its hidden state s_i at time i.

$$s_i = \text{LSTM}(s_{i-1}, y_{i-1}, c_i), \tag{4}$$

where c_i is distinct for each word y_i.

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.

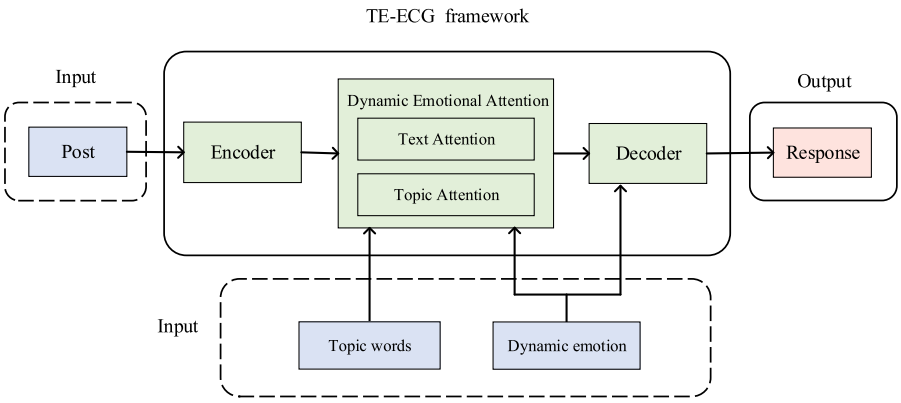

Fig. 1. Overview of TE-ECG. The green blocks form the TE-ECG framework, with the blue blocks as the input model, including the emotion and topic factors. The pink block represents the output from TE-ECG. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, the subsequent token y_i is generated by sampling from the output probability distribution p_i, denoted by

$$y_i \sim p_i = p(y_i|\{y_1, \ldots, y_{i-1}\}, c_i) \tag{5}$$

$$= softmax(w_p s_i), \tag{6}$$

where w_p is the weight parameter of the model.

3.2. Topic module

To generate a coherent response to the content of a given post, we consider employing some useful information into responses. Inspired by [11], we attempt to obtain topic words of a post and generate the responses by using topic words that can essentially reflect its related content. This idea is based on daily human conversation, where people often provide their responses associated with some related topic information of the message. Here, we introduce a Twitter latent Dirichlet allocation (LDA), which represents the state-of-the-art topic model for short texts [35], to acquire topic words of a post. Because we consider that the dataset of our model is collected from Weibo posts and replies that are as short as that in Twitter, we neither use the standard LDA, which might not work well, nor the author–topic [36] model, ignoring a single microblog that often belongs to a single topic [35]. The basic idea of a Twitter LDA model is that each text is uniquely assigned to a topic and that each word in the text is marked as a background or topic word of the assigned topic. The details can be found in [35].

First, we adopt the Gibbs sampling algorithm [35] to estimate the parameters of a Twitter LDA, trained using the training dataset. Then, we identify the topic of each post and the topic words of each topic from the pre-trained Twitter LDA model. The number of topics is set to 100 during pre-training. Finally, we assign a topic to each post, picking the top n words (i.e., $n = 30$ in our experiment) that have the highest probabilities but are not in the high-frequency vocabulary for each topic.

3.3. Dynamic emotional attention

General attention mechanism cannot capture the emotion information of texts. Thus, we present a dynamic emotional attention mechanism to incorporate emotion information into attention mechanism, inspired by [1]. This mechanism enables the model to focus dynamically on emotion-related information of the input text and topic words during decoding. Particularly, we use a dynamic emotion state to store emotional information that is similar to the internal memory presented in [1]. Some amount of the dynamic emotion state will be decayed during the decoding process. In this study, we further apply the dynamic emotion state to learn emotion-aware attention vectors from both the input texts and topic words, leading to the capture of emotion-related topic and input text information.

As shown in Fig. 2, we simultaneously feed the encoder's output $(h_1, h_2, \ldots, h_T)$ into the text attention and the embeddings of topic words $(t_1, t_2, \ldots, t_{30})$ into the topic attention. Then, the dynamic emotion state d_{i-1} is concatenated with the hidden state s_{i-1} to simultaneously compute topic and text attention. The text and topic attention are combined into dynamic emotional attention, which plays an essential role in obtaining important information of the post and emotion-related topic words for the response.

During decoding, at step i, the context vector c_i is acquired from Eq. (2), where the weight α_{ij} of h_j is updated as

$$\alpha_{ij} = \frac{\exp(\eta(s_{i-1}, d_{i-1}, h_j))}{\sum_{k=1}^{T} \exp(\eta(s_{i-1}, d_{i-1}, h_k))}. \tag{7}$$

Meanwhile, embeddings of topic words $(t_1, t_2, \ldots, t_{30})$ are linearly combined as a topic embedding o_i by topic attention, which is calculated as

$$o_i = \sum_{j=1}^{30} \alpha'_{ij} t_j, \tag{8}$$

where the combination weight α'_{ij} of t_j (similar to α_{ij} in Eq. (7)) is expressed as

$$\alpha'_{ij} = \frac{exp(\eta_o(s_{i-1}, d_{i-1}, t_j))}{\sum_{k=1}^{30} exp(\eta_o(s_{i-1}, d_{i-1}, t_k))}. \tag{9}$$

Consequently, emotion-related topic vectors $(o_1, o_2, \ldots, o_N)$ can obtain additional affective and diverse information regarding the content of a post. Then, emotion-related context vectors c_i and topic vector o_i are concatenated as the decoder's input, which contains the factors of specified emotion categories and additional information regarding a specific post.

We use the emotion category embedding, which can endow the model with emotional intelligence, as the initial state of the dynamic emotion state. Specifically, each emotion category is represented as a one-hot vector e and converted to a dense continuous category embedding by using a feed-forward network. At time step i, we compute a read gate g_r, which is used to read the dynamic emotion state d_{i-1}. The read gate is denoted by

$$g_r = sigmoid(w_r[s_{i-1}, y_{i-1}, c_i, o_i]), \tag{10}$$

where s_{i-1} is the hidden state of the previous time, y_{i-1} is the embedding of a previously decoded word, and c_i and o_i are the attention vectors of the input context and topic information, respectively.

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.

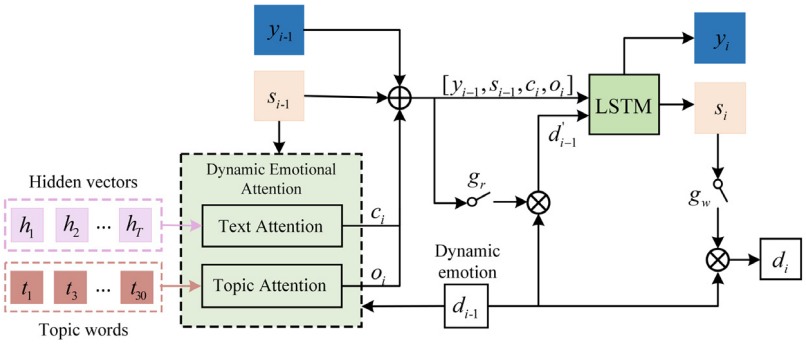

Fig. 2. Dynamic emotional attention. At time step i, c_i and o_i, which are from text and topic attention, respectively, are concatenated as the input of the decoder.

By reading the dynamic emotion state d_{i-1}, we can obtain a hidden emotion state d'_{i-1}, described as

$$d'_{i-1} = g_r \cdot d_{i-1}. \tag{11}$$

LSTM updates its hidden state s_i conditioned on the hidden state of the previous time s_{i-1}, previously decoded word embedding y_{i-1}, emotion-related context vector c_i and topic vector o_i, and hidden emotion state d'_{i-1} as follows:

$$s_i = \text{LSTM}(s_{i-1}, y_{i-1}, c_i, o_i, d'_{i-1}). \tag{12}$$

Finally, we update the dynamic emotion state d_i, formulated as

$$d_i = g_w \cdot d_{i-1}, \tag{13}$$

where g_w is a write gate, to attenuate the emotion state, computed as

$$g_w = sigmoid(w_w s_i). \tag{14}$$

4. Experiments

In this section, we present the details of our experiments. First, we introduce the dataset and parameter setting, and then describe our baselines and evaluation metrics. Then, we provide the detail of the experimental results and show several samples for intuitive comparison. To facilitate other researchers to perform the experiments, we will share the source codes with the research community.

4.1. Dataset

We use the dataset from ECG task released by The Sixth Conference on Natural Language Processing and Chinese Computing (NLPCC 2017) [32]. This dataset is obtained from Weibo posts and replies with more than 1 million post–response pairs. Note that each post and response pair of this dataset is equipped with an emotion label corresponding to an emotion category. As stated in [1,33], the result of emotion classification is obtained using the BiLSTM model, which is trained on the data from the NLPCC 2013 (http://tcci.ccf.org.cn/conference/2013) and 2014 (http://tcci.ccf.org.cn/conference/2014) emotion classification challenge.

During the preprocessing phase, we clean the origin dataset by removing excess punctuation and Cantonese text. To alleviate the out-of-vocabulary (OOV) problem, we construct a more fine-grained vocabulary composed of both high-frequency words and Chinese characters. First, we perform word frequency statistics on the dialogue data and select the words with a frequency of more than 80 to form a initial vocabulary. Then, we split the OOV into characters and added them to the initial vocabulary. Finally, the vocabulary contains 12,190 items, consisting of words and characters with a frequency of more than 80 and a UNK label and an END token. Additionally, we filter out the extremely short and meaningless post–response pairs to obtain better training data. The length of each post and response is limited to 20. We finally retain approximately 85% of the raw training data.

4.2. Parameter setting

Our proposed model is implemented on the Tensorflow platform (https://github.com/tensorflow/tensorflow). With 1000 hidden cells for each LSTM, the encoder uses a BiLSTM structure and the decoder is based on an LSTM structure. The word embeddings in our model are pre-trained using word2vec [37] with a vector size of 1000. The OOV words are represented as a zero vector. The vocabulary size is set to 12,819. Each emotion category is initialized as a one-hot vector of size six. The embedding size of the emotion category vector is set to 100 after the training stage. Finally, we set the number of topics to 100 and choose the top 30 topic words corresponding to each topic.

We optimize the objective function by using Adam, which has the ability of AdaGrad to handle sparse gradients and the capability of RMSProp to cope with non-stationary objectives [38]. The learning rate is set to 0.001 and the model is trained in a mini-batch of size 128. The maximum length of the posts and responses is limited to 20. Any sequence of length less or more than 20 is padded or cut, respectively, to unify the input length.

4.3. Baselines

Here, we consider two suitable baselines: a general seq2seq model with an attention mechanism (Seq2SeqA) [34] and an ECM [1], which is also proposed to integrate the emotional factor into a large-scale conversation generation system and can simultaneously generate emotional responses with various emotion categories. The setup of Seq2SeqA is the same as our model. However, no source code of the ECM is released; hence, we generally implement the ECM model according to [1] by using our dataset. Given that reproducing empirical results is challenging, the results in our study may disaccord with the published results of ECM to some extent.

A comparison between Seq2SeqA and TE-ECG shows that our model can generate responses that are not only varied in content but also appropriate in emotion. Since TE-ECG can consider emotion, we want to verify whether TE-ECG also does or does not ignore the quality and content relevance of the utterances compared to ECM. The first three tables below summarize the experimental results. The details are discussed in the following section.

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.

Table 2
Content and emotion scores of human evaluation on the various emotional types.

Models	Like		Sad		Disgust		Angry		Happy		Overall	
	Cont.	Emo.	Cont.	Emo.	Cont.	Emo.	Cont.	Emo.	Cont.	Emo.	Cont.	Emo.
ECM	1.44	0.74	1.27	0.55	1.31	0.6	1.03	0.46	1.25	0.51	1.26	0.57
TE-ECG	1.42	0.76	1.45	0.66	1.34	0.47	1.24	0.51	1.16	0.53	**1.32**	**0.59**

4.4. Evaluation metrics

We consider two types of evaluation metrics: manual and automatic. The manual evaluation of our model is based on the metrics of the ECM study. Furthermore, we invite three postgraduates, who have knowledge in NLP, with rich Weibo experience to evaluate the generated responses. The adopted metrics have two aspects. First, Content (i.e., Cont.) evaluates the grammar accuracy and estimates whether the logic and content of a response are appropriate. Second, Emotion (i.e., Emo.) is used to judge whether the emotional category of the generated response is consistent with the user-specified emotion.

The generated reply is marked by a uniform standard. The rating scale of the content is 0, 1, and 2. If the reply is grammatically accurate and coherent, then the content score is 2. If only the first condition is met (i.e., the responses are too universal, such as "Me too,"), the content score of the reply is 1. If none of the two conditions is satisfied (i.e., the response is not a complete sentence), then the content score is 0. The rating scale of emotion is 0 and 1. If the reply is consistent with the given emotion, then the score is 1; otherwise, the emotion score is 0.

We also consider automatic metrics that compare the generated responses with ground-truth responses. The bilingual evaluation understudy (BLEU) algorithm [39], which was widely adopted in MT, is not an appropriate evaluation criterion for conversation generation owing to its low correlation with manual evaluation, according to [40]. Thus, we choose distinct unigrams (i.e., distinct-1) and bigrams (i.e., distinct-2) to evaluate how informative and diverse the responses are, as proposed in [41]. We first calculate the numbers of different unigrams and bigrams in the responses and then divide the numbers by the total numbers of unigrams and bigrams. Higher ratios of distinct-1 and distinct-2 indicate more content information in the response.

4.5. Experimental results

Human evaluation

Two hundred posts are randomly selected from 5,000 test data to be used as the final manual evaluation set. We generate 1,200 responses for both ECM and TE-ECG, corresponding to six kinds of emotions. Then, annotators judge the quality of all responses in accordance with the human evaluation metrics. The annotation results of the first two tables below are based on these responses.

Table 2 shows the scores for each emotion category of the two models. As presented in the table, the overall performance of the TE-ECG model outperforms the ECM (two-tailed t-test, $p < 0.005$ for content). After the emotion-related topic module is incorporated, the performance of TE-ECG in content improves compared to ECM, because TE-ECG has a higher possibility to generate emotion-related topic words, such as "delicious," when the current post contains "lunch," because both these words are in the assigned topic of the current post. Moreover, TE-ECG is comparable with ECM with regard to emotion, indicating that the combined topic and emotional factors contribute to the generation of emotional responses. In addition, we use Fleiss' kappa [42] to evaluate the agreements among annotators. Fleiss' kappa values for content and emotion are 0.545 and 0.530, respectively, both indicating " Moderate agreement".

Table 3
Human annotation results in content and emotion.

Models	Content			Emotion	
	+2	+1	0	+1	0
ECM	41.8%	42.5%	15.7%	57.2%	42.8%
TE-ECG	**43.6%**	44.9%	11.5%	**58.4%**	41.6%

Table 4
Pairwise preferences of the three systems.

Preference	Seq2SeqA	ECM	TE-ECG
Seq2SeqA	–	40.80%	31.80%
ECM	59.20%	–	41.60%
TE-ECG	**68.20**%	**58.40**%	–

Table 5
Performance of our model and comparators on automatic evaluation of distinct-1 and distinct-2.

Metrics	Seq2SeqA	Models	Like	Sad	Disgust	Angry	Happy	Overall
distinct-1	0.055	ECM	0.064	0.042	0.044	0.042	0.056	0.05
		TE-ECG	0.076	0.051	0.053	0.048	0.070	**0.06**
distinct-2	0.184	ECM	0.225	0.142	0.181	0.146	0.172	0.173
		TE-ECG	0.255	0.171	0.218	0.173	0.212	**0.206**

The percentage of responses with different scores in content and emotion via human evaluation is shown in Table 3. As presented in the table, TE-ECG has a proportion of responses labelled "+1" of 58.4% in emotion, which is higher than 57.2% for ECM. Hence, our model can generate comparable emotion-suitable responses with ECM. Moreover, for content, 43.6% of the responses generated by TE-ECG are labelled as "+2", higher than 41.8% for ECM. TE-ECG increases by nearly 4% in "+2" responses, indicating that the emotion-related topic module indeed contributes to improving the quality of generated responses. This is because TE-ECG integrates the topic information into the generation of responses. Thus, the sentences are much more related in content and have slight possibility of being grammatically inaccurate.

To compare these models intuitively, the annotators are asked to select their preferred response to the same post in pairs. As shown in Table 4, TE-ECG preforms better than the other methods, indicating that TE-ECG is more popular with the annotators. With the combination of emotion-related topic and emotional factors, TE-ECG can generate more emotional and diverse responses that are more likely to be preferred.

Metric-based evaluation

Based on the automatic metrics, we evaluate the generated responses in content level. As mentioned previously, we utilize a general seq2seq model with an attention mechanism (Seq2SeqA) [34] and ECM as our baselines for comparison with our TE-ECG model.

We generate responses on the 5,000 test data by using baselines, as described in Section 4.3. The automatic evaluation results are presented in Table 5. As shown in the table, according to the automatic metrics, TE- ECG achieves superior performance. Compared with ECM, the performance of our model on both distinct-1 and distinct-2 improves significantly in every emotion category, and our model exhibits a 20% jump in unigram diversity (distinct-1) and a 19% increase in bigram diversity (distinct-2). Additionally,

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.

Table 6
Examples generated by each model (Chinese and corresponding English translation). The words marked in red represent the emotion words and informative words from the TE-ECG are underlined.

Post	Model		Response
这是哪里，这么美。 Where is this place, so beautiful.	Seq2SeqA		杭州，很美！ Hangzhou, very beautiful!
	ECM	Like	中山公园，很美，很美！ Zhongshan park, so beautiful.
		Sad	中山公园的路上，可惜我没去过 On the way to Zhongshan park. Unfortunately, I have not been to.
		Disgust	这是中山公园的，你说的是什么？你说的是什么？ This is in Zhongshan park. What are you talking about?
		Angry	这是中山的么?? Is this Zhongshan??
		Happy	哈哈，是的，美院。 Haha, yes, the Academy of Fine Arts.
	TE-ECG	Like	西藏 的一个小镇，很美 A small town in Tibet, so beautiful.
		Sad	深圳 啊，我都没去过 Shenzhen, I have not been to
		Disgust	深圳 啊，你也要去啊，你是说你的头像 吗? Shenzhen, you want to go, too? Are you talking about your avatar?
		Angry	深圳 啊!! 我在学校 啊! Shenzhen !! I am in the school!
		Happy	哈哈，在北京郊区 啊 Haha, in the suburb of Beijing
什么情况?? What happened?	Seq2SeqA		哈哈，你猜呢 Haha, you can guess
	ECM	Like	很可爱的小兔兔 Very cute little bunny.
		Sad	我觉得我很倒霉 I think I'm very unlucky
		Disgust	你说呢？你说呢？还是说你呢? What do you want to say? Talking about you?
		Angry	你说我是谁! You say, who I am!
		Happy	哈哈，你懂的 Haha, you know it.
	TE-ECG	Like	就是觉得很神奇，很神奇 I feel amazing
		Sad	我也不知道，莫名地失落 I don't know, I just feel upset
		Disgust	就是不晓得嘛，反正就是不晓得 I don't know, I don't know anyway.
		Angry	我就是个不明真相的群众 I am a mass who does not know the truth!
		Happy	哈哈，就是不知道啊 Haha, I don't know

note that our model's distinct-1 and distinct-2 scores are slightly higher than that of Seq2SeqA, probably because our model has to simultaneously consider the specified emotion and the diversity of content when generating a response. The results in Table 5 indicate that after combining emotion-related topic information with emotional factors, our system is more likely to ignore common words but prioritizes informative words in comparison with the responses generated by the baseline models.

Results comparison

Table 6 contains several sample responses when comparing TE-ECG with Seq2SeqA and ECM for additional intuitive understanding. The coloured words are emotion words, and informative words from the TE-ECG are underlined. The TE-ECG responses obtained using emotion and topic modules are more diverse and emotional compared to the responses generated by Seq2SeqA. In comparison with ECM, our model can generate higher-quality responses, which are more relevant to the corresponding posts. The experimental results show intuitively that our model can address the problem of conversation generation at emotional and content levels to some extent.

5. Conclusion and future work

This paper presents a TE-ECG, consisting of topic and dynamic emotional attention modules. The topic module aimed to identify a topic to ensure that the response and the corresponding post have particular correlations. The model used topic words of the identified topic as extra prior knowledge to ensure that content relevance for ECG is not neglected. Moreover, by using the dynamic

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.

emotional attention, the system can combine the content with emotional factors in the decoding process to generate utterances that are not only diverse and informative, but also related to a specified emotion. We conducted the experiments on large-scale Weibo post–response pairs released by the NLPCC 2017 ECG task. The experimental results showed that our model achieves good performance, even outperforming the popular seq2seq conversation generation models.

We plan to continue this work in the following directions. First, the current system is still unsteady in generating more diverse and intellective responses in daily conversation. A natural avenue for future research should enrich the emotion and content of the responses based on external knowledge. Second, since the generated emotional responses occasionally occur in a few repeated and dull expressions, a more effective algorithm that can generate few repeated words and general expressions must be developed in the future. Furthermore, we will extend our experiments on additional language datasets.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 61573163) and the Wuhan Youth Science and Technology plan.

References

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

[2] L. Shang, Z. Lu, H. Li, 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 of the Asian Federation of Natural Language Processing, Beijing, China, 2015, pp. 1577–1586.

[3] H. Wang, Z. Lu, H. Li, E. Chen, A dataset for research on short-text conversations, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, 2013, pp. 935–945.

[4] Z. Ji, Z. Lu, H. Li, An information retrieval approach to short text conversation, 2014. ArXiv preprint arXiv:1408.6988.

[5] E. André, M. Rehm, W. Minker, D. Bühler, Endowing spoken language dialogue systems with emotional intelligence, in: Affective Dialogue Systems, Tutorial and Research Workshop, Kloster Irsee, Germany, 2004, pp. 178–187.

[6] M. Skowron, Affect listeners: acquisition of affective states by means of conversational systems, in: Development of Multimodal Interfaces: Active Listening and Synchrony, Second COST 2102 International Training School, Dublin, Ireland, 2009, pp. 169–181.

[7] J.D. Williams, S.J. Young, Partially observable markov decision processes for spoken dialog systems, Comput. Speech Lang. 21 (2) (2007) 393–422.

[8] D. Bohus, A.I. Rudnicky, A principled approach for rejection threshold optimization in spoken dialog systems, in: 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, 2005, pp. 2781–2784.

[9] A. Ritter, C. Cherry, W.B. Dolan, Data-Driven response generation in social media, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 2011, pp. 583–593.

[10] L. Mou, Y. Song, R. Yan, G. Li, L. Zhang, Z. Jin, Sequence to backward and forward sequences: a Content-Introducing Approach to Generative Short-Text Conversation, in: 26th International Conference on Computational Linguistics, Osaka, Japan, 2016, pp. 3349–3358.

[11] C. Xing, W. Wu, Y. Wu, J. Liu, Y. Huang, M. Zhou, W.-Y. Ma, Topic aware neural response generation, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017, pp. 3351–3357.

[12] N. Asghar, P. Poupart, J. Hoey, X. Jiang, L. Mou, Affective neural response generation, 2017. ArXiv preprint arXiv:1709.03968.

[13] M. Ghazvininejad, C. Brockett, M.W. Chang, B. Dolan, J. Gao, W.t. Yih, M. Galley, A knowledge-grounded neural conversation model, 2017. ArXiv preprint arXiv:1702.01932.

[14] I.V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A.C. Courville, Y. Bengio, A hierarchical latent variable encoder-decoder model for generating dialogues, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017, pp. 3295–3301.

[15] I.V. Serban, A. Sordoni, Y. Bengio, A.C. Courville, J. Pineau, Building end-to-end dialogue systems using generative hierarchical neural network models, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA, 2016, pp. 3776–3784.

[16] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, in: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, Montreal, Quebec, Canada, 2014, pp. 3104–3112.

[17] A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J. Nie, J. Gao, B. Dolan, A neural network approach to context-sensitive generation of conversational responses, in: The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, 2015, pp. 196–205.

[18] R. Yan, D. Zhao, W. E., Joint learning of response ranking and next utterance suggestion in human-computer conversation system, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, ACM, 2017, pp. 685–694.

[19] J. Li, M. Galley, C. Brockett, G.P. Spithourakis, J. Gao, W.B. Dolan, A persona-based neural conversation model, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 2016.

[20] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, D. Jurafsky, Adversarial learning for neural dialogue generation, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017, pp. 2157–2169.

[21] J. Li, W. Monroe, D. Jurafsky, A simple, fast diverse decoding algorithm for neural generation, 2016. ArXiv preprint arXiv:1611.08562.

[22] A.K. Vijayakumar, M. Cogswell, R.R. Selvaraju, Q. Sun, S. Lee, D. Crandall, D. Batra, Diverse beam search: decoding diverse solutions from neural sequence models, 2016. ArXiv preprint arXiv:1610.02424.

[23] L. Shao, S. Gouws, D. Britz, A. Goldie, B. Strope, R. Kurzweil, Generating long and diverse responses with neural conversation models, 2017. ArXiv preprint arXiv:1701.03185.

[24] M. Skowron, S. Rank, M. Theunis, J. Sienkiewicz, The good, the bad and the neutral: affective profile in dialog system-user communication, in: Affective Computing and Intelligent Interaction — 4th International Conference, Memphis, TN, USA, 2011, pp. 337–346.

[25] M. Ptaszynski, P. Dybala, W. Shi, R. Rzepka, K. Araki, Towards context aware emotional intelligence in machines: computing contextual appropriateness of affective states, in: Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, 2009, pp. 1469–1474.

[26] X. Sun, X. Peng, S. Ding, Emotional human-machine conversation generation based on long short-term memory, Cogn. Comput. (2017) 1–9.

[27] R. Zhang, Z. Wang, D. Mai, Building emotional conversation systems using multi-task seq2seq learning, in: Natural Language Understanding and Intelligent Applications — 5th CCF Conference on Natural Language Processing and Chinese Computing, and 24th International Conference on Computer Processing of Oriental Languages, Dalian, China, 2017.

[28] Y. Zhuang, X. Wang, H. Zhang, J. Xie, X. Zhu, An ensemble approach to conversation generation, in: Natural Language Understanding and Intelligent Applications — 5th CCF Conference on Natural Language Processing and Chinese Computing, and 24th International Conference on Computer Processing of Oriental Languages, Dalian, China, 2017.

[29] J. Xuan, X. Luo, G. Zhang, J. Lu, Z. Xu, Uncertainty analysis for the keyword system of web events, IEEE Trans. Syst. Man Cybernet. Syst. 46 (6) (2016) 829–842.

[30] Y. Rao, Q. Li, X. Mao, L. Wenyin, Sentiment topic models for social emotion mining, Inf. Sci. 266 (2014) 90–100.

[31] Y. Rao, Q. Li, W. Liu, Q. Wu, X. Quan, Affective topic model for social emotion detection, Neural Netw. 58 (2014) 29–37.

[32] X. Huang, J. Jiang, D. Zhao, Y. Feng, Y. Hong (Eds.), Natural Language Processing and Chinese Computing — 6th CCF International Conference, NLPCC 2017, Dalian, China, November 8–12, 2017, Proceedings, in: Lecture Notes in Computer Science, vol. 10619, Springer, 2018.

[33] M. Huang, Z. Ye, H. Zhou, Overview of the nlpcc 2017 shared task: emotion generation challenge, in: National CCF Conference on Natural Language Processing and Chinese Computing, 2017, pp. 926–936.

[34] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014. ArXiv preprint arXiv:1409.0473.

[35] W.X. Zhao, J. Jiang, J. Weng, J. He, E. Lim, H. Yan, X. Li, Comparing twitter and traditional media using topic models, in: Advances in Information Retrieval — 33rd European Conference on IR Research, Dublin, Ireland, 2011, pp. 338–349.

[36] M. Steyvers, P. Smyth, M. Rosen-Zvi, T.L. Griffiths, Probabilistic author-topic models for information discovery, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, 2004, pp. 306–315.

[37] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013, Lake Tahoe, Nevada, United States, 2013, pp. 3111–3119.

[38] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: International Conference on Learning Representations, San Diego, California, USA, 2015.

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.

[39] K. Papineni, S. Roukos, T. Ward, W. Zhu, Bleu: a method for automatic evaluation of machine translation, in: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 2002, pp. 311–318.

[40] C. Liu, R. Lowe, I. Serban, M. Noseworthy, L. Charlin, J. Pineau, 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, Austin, Texas, USA, 2016, pp. 2122–2132.

[41] J. Li, M. Galley, C. Brockett, J. Gao, B. Dolan, A diversity-promoting objective function for neural conversation models, in: The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, 2016, pp. 110–119.

[42] J.L. Fleiss, Measuring nominal scale agreement among many raters., Psychol. Bull. 76 (5) (1971) 378.

Please cite this article in press as: Y. Peng, et al., Topic-enhanced emotional conversation generation with attention mechanism, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.09.006.