SSM-Seq2Seq: A Novel Speaking Style Neural Conversation Model

Boran Wang¹* and Yingxiang Sun¹
¹Beijing Institute of Technology, Beijing, China
*Email: wangbr@bit.edu.cn

Abstract. Open domain personalized dialogue system has attracted more and more attention because of the ability of generating interesting and personalized responses. To incorporate speaking style, the existing methods first train respectively a response generator over a non-personalized conversational dataset and a speaking style extractor over a personalized non-conversational dataset, and then generate personalized responses by the parameter sharing mechanism. However, the training datasets' speaking styles of the response generator and speaking style extractor is totally different, which makes the performance of the existing methods be not optimal. Intuitively, it will improve the performance by decreasing the gap between two training datasets' speaking styles. Thus, in this paper, we propose a novel speaking style memory sequence-to-sequence (SSM-Seq2Seq) model, which incorporates the speaking style information from personalized non-conversational dataset into the training dataset of response generator to eliminate the gap. Extensive experiments show that the proposed approach yields great improvement over competitive baselines.

1. Introduction
Personalized dialogue system has attracted much attention for its huge and attractive commercial value. Existing personalized dialogue systems can be classified into two categories, task-oriented personalized dialogue systems and open domain personalized dialogue system. Task-oriented personalized dialogue system [1-2] aim to assist the user to complete certain tasks (e.g., reserving accommodations and restaurants) better and faster than non-personalized dialogue systems according to the preferences and habits of a user. Open domain personalized dialogue systems [3-5] make users feel they are talking to a particular real person. For example, if a sick old man wants to talk to his son, a personalized dialogue system can be given the personality of his son, which can better solve the problem of emotional support. In this paper, we focus on the open domain personalized dialogue systems.
There are mainly two kinds of methods for building an open domain personalized dialogue system. The first category is to train a Seq2Seq model on normal conversational dataset [5]. Then it is fine-tuned on the personalized conversational corpus that could reflect the speaking style of the target person. However, this kind of specific dataset need to be constructed manually, the cost is high. In order to alleviate this problem, the other category incorporates speaking style from non-conversational dataset called stylized texts, such as blog posts, into dialogue system [4,6]. As illustrated in Figure 1, this kind of methods consists of three components: response generator, speaking style extractor and parameter sharing mechanism. The response generator is responsible for generating fluent and reasonable responses training on non-personalized conversational dataset. The speaking style extractor extracts the speaking style information from stylized texts. Finally, the framework uses the parameter sharing mechanism to generate fluent, reasonable and stylized responses at the same time.

However, the training datasets' speaking style of response generator and speaking style extractor exists the gap. The gap may weaken the expression of speaking style information in the final generated responses. Generally, the two components are trained on non-personalized conversational dataset and personalized conversational dataset, respectively. The speaking style information is different in the two datasets. Because of the parameter sharing mechanism, the gap may weaken the model's ability to generate stylized responses. Intuitively, if we make the training datasets' of the response generator and the speaking style extractor contain similar speaking style information, the final generated responses will be influenced greatly by the response generator and have more speaking style information. Thus, we incorporate the speaking style information into the response generator and combine the dialogue context information with speaking style information before the parameter sharing mechanism.

According to this main idea, we propose a novel speaking style memory sequence-to-sequence (SSM-Seq2Seq) model to generate more stylized responses for different scenarios. On the one hand, we train a standard Seq2Seq module that encodes conversation history combined with style speaking style information and generate responses. The speaking style information is stored and modeled by memory network. On the other hand, an autoencoder [7,8] module is trained to predict the input sequence itself. The input sequence is a sentence in the stylized texts. And the two modules share the parameters of the decoder part.

Our major contributions are as follows:

- We propose a novel speaking style memory sequence-to-sequence (SSM-Seq2Seq) model, which incorporate the speaking style information from stylized texts into the training dataset of response generator to eliminate the gap between two components' training datasets, which is beneficial to generate more stylized responses.
- We propose an efficient automatic metric, which can evaluate the degree of speaking style in the generated responses.
- Extensive experiments show that the proposed SSM-Seq2Seq can generate more interesting and personalized responses, and outperform state-of-the-art response generation models.

![Figure 1. Three components in the second kind of methods.](image)
2. Related Work
There has been a large amount of work for conversation generation. [9-11] used Seq2Seq model that are efficient in the translation task to train an end-to-end dialogue system. Seq2Seq model utilizes neural networks to represent dialogue history and generate appropriate responses, while requiring very little domain knowledge and manual annotation.

In recent years, a lot of works have focused on incorporating personal information into the open-domain conversation model. [12] encoded personas in distributed embeddings that capture individual characteristics such as background information and speaking style. [13] presented a model that can generate responses conforming to a pre-specified agent profile. And instead of learning personality from dialogue data, them work can assign a desired identity to a chatbot. [5] leveraged target conversation data about speakers’ personal information, such as age and gender, to condition generation using domain adaptation methods. [3] introduced the PERSONA-CHAT dataset consisting of paired human generated profiles and conversations. They leverage this data to construct the agents that have consistent personalities. However, the works all need special conversation data reflecting speakers’ personal information. Personal conversational data is scarce. Considering this problem, [6] proposed three simple yet effective methods of incorporate the speaking style from a small monologue speaker corpus. Three methods in the neural encoder-decoder framework. [4] introduced a multi-task learning and parameter sharing mechanism to incorporate speaker role characteristics into conversational models using personalized non-conversational dataset. However, speaking style information of personalized non-conversational dataset and non-personalized conversational dataset is totally different. The difference weakens the expression of speaking style information in the final generated responses because of parameter sharing mechanism. In order to eliminate this difference, we extract speaking style information from stylized text and add it to the conversational dataset, which could enable model to generate more stylized responses.

3. The proposed model
Our problem can be formulated as follows: Given a conversation input \( X = (x_1, x_2, \ldots, x_n) \) and some stylized texts \( S = (s_1, s_2, \ldots, s_k) \) that could reflect the speaking style of the target speaker. The goal is to generate a proper response \( Y = (y_1, y_2, \ldots, y_m) \), which not only is appropriate for conversation scene but also could reflect speaking style of the target speaker. Essentially, the model estimates the probability:

\[
P = (Y|X, S) = \prod_{t=1}^{m} P(y_t|y_{<t}, X, S)
\]

Figure 2. SSM-Seq2Seq model architecture.
The overview of the proposed conversational model is presented in Figure 2. The model can be divided into four parts: Seq2Seq, speaking style encoder, autoencoder and parameter sharing mechanism. In the Seq2Seq module, both conversation input and contextually relevant stylized texts are fed into two distinct encoders. One of the two encoders is speaking style encoder to encode contextually relevant stylized texts. The other one is responsible for encoding dialogue context information. Decoder is used to predicts the responses based on the output of two encoders. The autoencoder is used to predicts the input sequence itself and is trained on stylized texts. Finally, we share the decoder parameters to generate stylized responses. We will give more details about each part in the following sections.

3.1. Seq2Seq module

Inspired by the effectiveness of Seq2Seq models in neural machine translation task [14-15], researchers have been adopting these techniques to implement response generation model and have achieved significant improvements [9-11]. Given a sequence of inputs \(X = (x_1, x_2, \ldots, x_n)\), Seq2Seq models use a Long Short-Term Memories (LSTM) [16] to encode a variable-length input sequence and generates hidden states \(H = (h_1, h_2, \ldots, h_n)\) as the representation of \(X\). The decoder that is another LSTM is initialized by \(H_n\) and generates another sentence \(Y = (y_1, y_2, \ldots, y_n)\) as an output. Each sentence terminates with a special end-of-sentence symbol EOS. During decoding, the algorithm terminates when an EOS token is predicted. This part of our model is almost identical to prior conversational Seq2Seq models, except that we encode the extra speaking style information in the encoding phrase.

3.2. Speaking Style Encoder

In our adaptation of memory networks, we firstly select 20 sentences from stylized texts. And they have the highest semantic similarity with Seq2Seq module input, namely, conversation context. Specifically, we represent a sentence by averaging the weighted pre-trained word embeddings [17] of all words and the weight of a word is given by TF-IDF [18]. We compute the cosine score between conversation input and stylized texts. The 20 sentences with the highest score are chosen. We encode them considering the similarity between stylized texts and conversation input from word and sentence level. In the sentence level, we use a LSTM to encode each \(s_i \in S_r\) as hidden vectors \((h_1, h_2, \ldots, h_T)\) and the last hidden vector \(h_t\) is regarded as the vector representation of \(s_i\). We denote the vector representation as \(u\). \(u\) is a weighted summation of \(S_r\) defined by

\[
    u = \sum_{i=1}^{n} p_i s_i
\]

The weight \(p_i\) is calculated as

\[
    p_i = \frac{\exp(o^T s_i)}{\sum_k \exp(o^T s_k)}
\]

\(o\) is the vector representation of conversation input. \(r\) is the number of semantic relevant stylized sentences.

In the word level, firstly given the conversation input \((a_1, a_2, \ldots, a_n)\) and a semantic relevant stylized sentence \((b_1, b_2, \ldots, b_m)\) we have

\[
    e_{ij} = a_i^T b_j
\]

\[
    \hat{a}_i = \sum_{j=1}^{m} \frac{\exp(e_{ij})}{\sum_{k=1}^{m} \exp(e_{ik})} b_j, \forall i \in [1, \ldots, n]
\]

\[
    r = \frac{1}{n} \sum_{i=1}^{n} \hat{a}_i
\]

Where \(\hat{a}_i\) is a weighted summation of \(b_j\). Intuitively, the speaking style in \(b_j\) that is relevant to \(a_i\) will be selected and represented as \(\hat{a}_i\). And \(r\) is the a new representation of conversation input.
considering a stylized sentence. The same operation is performed for each stylized sentences in the $S_r$, repeatedly. So we can get $v$ defined by

$$v = \frac{1}{r} \sum_{i=1}^{r} r_i$$

(6)

Finally, the hidden state of the LSTM decoder is initialized with $\hat{\delta}$, which is a combination of input sequence and the relevant stylized texts, to predict the responses sentence word by word. And $\hat{\delta}$ is defined by

$$\hat{\delta} = o + v + u$$

(7)

We combine the conversation context information with the speaking style information by speaking style encoder. The mixed information serves as the input of the decoder which is sharing parameters between Seq2Seq module and autoencoder module. It reduces the gap so that can help generate more stylized responses.

3.3. Autoencoder module

In the autoencoder module, only stylized texts are to be used for training. Like the Seq2Seq module, an autoencoder module consists of encoder and decoder. However, unlike the source to target mapping in the Seq2Seq model, autoencoder is used to predicts the input sequence itself.

3.4. Parameter sharing mechanism

We only share the decoder parameters of Seq2Seq module and autoencoder module in the parameter sharing mechanism, so that the final generated responses can be guide which can help model generate stylized responses by incorporating the speaking style of the target person. The others parameters are not tied and are learned.

4. Experiment

4.1. Dataset

Conversation Dataset: We evaluate our model on a Chinese corpus released by [9]. This corpus consists of 4.44 million message-response conversation pairs obtained from Sina Weibo. Weibo is a popular Twitter-like microblogging service in China, on which a user can post short messages. Other users make comment on a published post, which will be referred to as responses. We divide the dataset into 3 sets, there are 20K/20K/3.50M pairs in the test/dev/training sets, respectively.

Conversation Dataset: In order to collect non-conversation data reflecting the speaking style of the target person, we select 5 different weibo user and crawl their released message. The users have different careers, ages, genders. They have totally different speaking styles.

4.2. Baselines

Seq2Seq: A Seq2Seq [15] model, which is widely used in open-domain conversational systems and was trained without non-conversation data.

Seq2Seq+MTASK: A Multi-task model based on Seq2Seq [4]. It trains a standard seq2seq model that encodes conversation history and generate responses. What’s more an autoencoder model is trained to predict speaking style texts given the same speaking style texts. The two models share the parameters of the decoder part.

Seq2Seq+Fine-tune: A fine-tuning model based on Seq2Seq [6]. The authors propose three simple methods of influencing the speaking style of the output in the neural encoder-decoder frame. We choose the best performance model of the three models as a baseline. This method treats the sentence in the stylized texts as a response sentence $R$, which is used to generate a pseudo input sentence $I$ based on the backward model $p(R|I)$. The backward model $p(I|R)$ is trained on non-
personalized conversational dataset. What's more, it also treats the previous sentence and the current sentence in the styled texts as a pseudo context and a response. Then this method uses the pseudo contexts and correspondent the response sentences in the stylized texts as the conversation pairs. The stylized conversation pairs are used to finetune the standard Seq2Seq model trained over the non-personalized conversational dataset.

4.3. Automatic Evaluation

**Metrics:** We chose several suitable existing automatic metrics and propose four new metrics to evaluate the model's ability of generating stylized responses.

- **BLEU:** BLEU [9-10] has been shown to correlate well with human judgment on the response generation task.
- **DISTINCT-1/2:** DISTINCT-1/2 proposed by [19] measures how diverse and informative the generated responses are. DISTINCT-1/2 is the ratio of distinct uni-grams/bigrams in generated responses.

### Table 1. Stylized vector similarities(SVS) between different persons.

| SVS    | User1-2 | User2-2 | User3-2 | User4-2 | User5-2 |
|--------|---------|---------|---------|---------|---------|
| User1-1| 0.9853  | 0.9069  | 0.9071  | 0.9104  | 0.8875  |
| User2-1| 0.8956  | 0.9922  | 0.9035  | 0.9148  | 0.8846  |
| User3-1| 0.9303  | 0.9101  | 0.9902  | 0.9536  | 0.8890  |
| User4-1| 0.8873  | 0.9023  | 0.9344  | 0.9795  | 0.8742  |
| User5-1| 0.7239  | 0.7644  | 0.7607  | 0.8621  | 0.9100  |

### Table 2. Words overlap between different persons.

| word overlap | User1-2 | User2-2 | User3-2 | User4-2 | User5-2 |
|--------------|---------|---------|---------|---------|---------|
| User1-1      | 0.4173  | 0.2419  | 0.2678  | 0.2415  | 0.2587  |
| User2-1      | 0.3621  | 0.4599  | 0.3312  | 0.2819  | 0.3195  |
| User3-1      | 0.2742  | 0.2225  | 0.5323  | 0.3507  | 0.2383  |
| User4-1      | 0.2119  | 0.1736  | 0.2388  | 0.3827  | 0.1902  |
| User5-1      | 0.2789  | 0.2282  | 0.2628  | 0.4177  | 0.5368  |

### Table 3. Classification accuracies.

| Target User | Accuracy |
|-------------|----------|
| User1       | 0.8471   |
| User2       | 0.8953   |
| User3       | 0.9204   |
| User4       | 0.8532   |
| User5       | 0.8418   |
| Normal      | 0.8735   |

- **Stylized Vector Similarity and Words Overlap:** We observed that word-using habit partly reflects speaking style. So we propose two methods called stylized vector similarity and words overlap, which measure the word-using habits of the person. For the first method, we select 100 words with the highest word frequency from stylized texts. We denote the words as $W = \{w_1, w_2, \ldots, w_{100}\}$. We build two vectors $V_r = \{v_{r1}, v_{r2}, \ldots, v_{r100}\}$ and $V_s = \{v_{s1}, v_{s2}, \ldots, v_{s100}\}$ for generated responses and stylized texts, respectively. $v_{r1}$ is the word frequency of $w_1$ in the generated responses. Similarly, $v_{s1}$ is the word frequency of $w_1$ in the stylized texts. Two vectors are divided by the total number of words in the generated responses and stylized texts, respectively. Finally, stylized vector similarity is computed using cosine similarity between two vectors. We do
not use all words frequency to constructing stylized vector. Because the influence of low frequency words for calculating vectors similarity is very small compared with high frequency words. In order to consider the similarity of usage of low frequency words between two texts, we propose another method called words overlap. We build two vocabularies for generated responses and stylized texts and count the number of words \( n_{both} \) in both vocabularies. \( n_{both} \) divided by the size of stylized text’s vocabulary could measure the the similarity of low frequency word between generated responses and stylized texts.

In order to verify the validity of the two methods, we divide the stylized texts from the same person into two parts in a ratio of 1:1. For stylized texts of five persons we can get ten stylized texts and we calculate stylized vector similarity and words overlap among ten stylized texts. As shown in Table 1 and Table 2, stylized vector similarity and words overlap of two stylized texts from the same person are quiet similar but the results of different persons are totally different.

- **Speaking Style Classifier:** The evaluation methods of stylized vector similarity and words overlap mainly consider the words-using habit between two speaking styles. So we propose another method called speaking style classifier that focuses on the speaking style of sentence level, such as: phrasing, personality. We use a bi-directional LSTM to encode the sentence and the last hidden states are then passed to a fully-connected softmax layer that outputs probabilities over six labels, namely stylized texts of five persons and non-personalized sentences from Weibo conversation dataset. In the evaluate phrase, we use this classifier to classify generated responses and treat the output probabilities of target person as speaking style score of the same person. To verify the ability of our speaking style classifier, we evaluate classification accuracy (i.e., the percentage of correctly labeled sentences) in six classifications. As shown in Table 3, the classifier can accurately distinguish the speaking styles of different persons.

**Result:** The evaluation results on existing and newly proposed automatic metrics for four models are shown in Table 4 and Table 5 respectively.

| Model       | BLEU-1 | BLEU-2 | BLEU-3 | BLEU-4 | Distinct-1  | Distinct-1  |
|-------------|--------|--------|--------|--------|-------------|-------------|
| Seq2Seq     | 0.1258 | 0.0579 | 0.0431 | 0.0358 | 0.0637      | 0.1849      |
| Seq2Seq+MTASK | 0.1282 | 0.0636 | 0.2542 | 0.0376 | 0.0728      | 0.2472      |
| Seq2Seq+Fine-tune | 0.1264 | 0.0613 | 0.0443 | 0.0372 | 0.0692      | 0.2215      |
| SSM-Seq2Seq  | **0.1304** | **0.0626** | **0.0454** | **0.0370** | **0.0821** | **0.2956** |

| Model       | SVS    | Words Overlap | Classifier |
|-------------|--------|---------------|------------|
| Seq2Seq     | 0.7235 | 0.0522        | 0.2371     |
| Seq2Seq+MTASK | 0.7861 | 0.0636        | 0.2542     |
| Seq2Seq+Fine-tune | 0.7649 | 0.0592        | 0.2513     |
| SSM-Seq2Seq  | **0.8134** | **0.0717**    | **0.2764** |

The proposed SSM-Seq2Seq outperforms baseline on most metrics. In comparison with the standard Seq2Seq model, SSM-Seq2Seq notably outperforms it in all metrics, which demonstrates SSM-Seq2Seq is able to generate a more appropriate and stylized response. What’s more compared with the Seq2Seq+MTASK model and Seq2Seq+Fine-tune model, we obtain an increase in DINSTINCT-1/2, which value is scaled by total number of generated tokens to avoid favoring long sentences. We interpret it means that our approach can help the system generate more diverse responses. Significant gains are obtained in three metrics of speaking style which show SSM-Seq2Seq
could incorporate speaking style into responses. However, SSM-Seq2Seq performance is approximation on BLUE-1, BLUE-2 and BLUE-3, but worse than Seq2Seq+MTASK and Seq2Seq+Fine-tune model on BLUE-4 score. A possible reason is that the generated responses of appropriateness and stylization are trade-offs. Because BLUE measures the similarity between standard responses and generated responses. The generated responses with speaking style likely have a deviation from standard responses. But we think the loss on BLUE scores is accepted considering the promotion of other metrics.

4.4. Human Evaluation

**Metrics:** We employ three native speakers to individually annotate 100 randomly generated responses from different models. With stylized texts of target person as example texts, judges are asked to select which system output appeared most likely to have been produced by the same person. Ties were permitted.

**Table 6.** Results of human evaluation after removing “Tie” pairs. SSM-Seq2Seq is significantly better than all the baselines on all the test sets.

| Model                      | SVS         |
|---------------------------|-------------|
| SSM-Seq2Seq vs Seq2Seq    | 0.5263 vs 0.4737 |
| SSM-Seq2Seq vs Seq2Seq+MTASK | 0.5125 vs 0.4875 |
| SSM-Seq2Seq vs Seq2Seq+Fine-tune | 0.5138 vs 0.4861 |

**Result:** Table 6 shows the results of human evaluation. SSM-Seq2Seq perform better on average than other models, consistent with the proposed automatic evaluation of speaking style results. The result indicates that our model could generate more stylized responses and newly proposed evaluation methods could effectively measure the degree of speaking style. There are some responses that could not be judged which is most likely to have been produced by the same person. Because similar responses from the same inputs may be insufficient for distinguishing different styles.

4.5. Case Study

We give a qualitative case study in Table 7. We expect that the responses could mimic the speaking style of the target speaker who is a kind and polite salesman. Responses generated by Seq2Seq are fluent. While this kind of response is appropriate for many other speakers, lacking speaking style. In the case 1 and case 3, the responses of our model expresses more sincere wishes and sympathy. In the case 2, SSM-Seq2Seq politely compliment the photograph taken by the other talker. By observing, it’s easy to find responses of SSM-Seq2Seq are more likely generated by a target speaker.

**Table 7.** Responses produced by our baseline and SSM-Seq2Seq response generation system, given a conversation input.

| Case | 1                        | 2                        | 3                        |
|------|--------------------------|--------------------------|--------------------------|
| Input| 希望姚明没事，早日康复，回到球场！ | 北京的天气真不错。 | 看着好可怜。 | 孩纸，西瓜跟你有仇啊！ | Why does this child treat watermelons like that? |
| Seq2Seq | I hope Yao Ming is all right, recover soon, and return to the court! | The weather in Beijing is very good these two days. Share photos of spring. | I particularly like these photos. | I particularly like these photos. | The pictures you take are very nice. I like them very much. |
| Seq2Seq+MTASK | 我也是。 So do I. | Is the weather good in Beijing? | 北京的天气真不错。 | 北京的天气真不错。 | The weather in Beijing is really good. |
| Seq2Seq+Fine-tune | 等着他们回来。Waiting for them to come back. | 北京的天气好？ | 怎么回事？ | what's the matter? | I feel painful. |
| SSM-Seq2Seq | 一起祈祷吧！Pray together! | 这必须点赞。 | 贫困地区的孩子吧，不觉得多搞笑，挺让人心疼的。 | 贫困地区的孩子吧，不觉得多搞笑，挺让人心疼的。 | These children come from poor areas. I don't think it's funny. It's very painful. |
|                | 真心希望大家都健康康，所有的伤痛都远离你们。 | 你拍的真的很好,我很喜欢。 | These children come from poor areas. I don't think it's funny. It's very painful. | These children come from poor areas. I don't think it's funny. It's very painful. | These children come from poor areas. I don't think it's funny. It's very painful. |
|                | I sincerely hope that all of you are healthy, all the injuries are away from you. | The pictures you take are very nice. I like them very much. | These children come from poor areas. I don't think it's funny. It's very painful. | These children come from poor areas. I don't think it's funny. It's very painful. | These children come from poor areas. I don't think it's funny. It's very painful. |
|                |                           |                           |                           |                           |                           |
5. Conclusion
In this paper, we proposed a novel speaking style memory sequence-to-sequence (SSM-Seq2Seq) model which leverages the generated parallel corpus to train a stylized response generation model. The parallel corpus integrates speaking style information from personalized non-conversational dataset with general answering pattern from the training dataset to facilitate stylized response. Automatic and human evaluation proves the validity of the proposed SSM-Seq2Seq.

Acknowledgments
We thank classmates and teachers for their helpful suggestions and discussions.

References
[1] Mo, K., Zhang, Y., Li, S., Li, J. and Yang, Q. 2018 Personalizing a dialogue system with transfer reinforcement learning. In: AAAI Conference on Artificial Intelligence. New Orleans. pp. 5317-5324.
[2] Oraby S, Reed L, Tandon S, Sharath TS, Lukin S, Walker M. 2018 Controlling Personality-Based Stylistic Variation with Neural Natural Language Generators. In: the 19th Annual SIGdial Meeting on Discourse and Dialogue. Stockholm. pp.180-190.
[3] Zhang S, Dinan E, Urbanek J, Szlam A, Kiela D, Weston J. 2018 Personalizing Dialogue Agents: I have a dog, do you have pets too? In: the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne. pp.2204-2213.
[4] Luan Y, Brockett C, Dolan B, Gao J, Galley M. 2017 Multi-Task Learning for Speaker-Role Adaptation in Neural Conversation Models. In: the Eighth International Joint Conference on Natural Language Processing. Taipei. pp.605-614.
[5] Zhang, W.N., Zhu, Q., Wang, Y., Zhao, Y. and Liu, T. 2019 Neural personalized response generation as domain adaptation. In: World Wide Web. San Francisco. pp.1427-1446.
[6] Wang D, Jojic N, Brockett C, Nyberg E. 2017 Steering Output Style and Topic in Neural Response Generation. In: Empirical Methods in Natural Language Processing. Copenhagen. pp. 2140-2150.
[7] Ng A. 2011 Sparse autoencoder. http://ailab.chonbuk.ac.kr/seminar_board/pds1_files/sparseAutoencoder.pdf
[8] Lu X, Tsao Y, Matsuda S, Hori C. 2013 Speech enhancement based on deep denoising autoencoder. In: Interspeech. Lyon. pp. 436-440.
[9] Shang L, Lu Z, Li H. 2015 Neural Responding Machine for Short-Text Conversation. In: the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing. pp. 1577-1586.
[10] Sordoni A, Galley M, Auli M, Brockett C, Ji Y, Mitchell M, Nie JY, Gao J, Dolan B. 2015 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. pp. 196-205.
[11] Vinyals O, Le Q. 2015 A neural conversation model. https://arxiv.org/abs/1506.05869.
[12] Li J, Galley M, Brockett C, Spithourakis G, Gao J, Dolan B. 2016 A Persona-Based Neural Conversation Model. In: the 54th Annual Meeting of the Association for Computational Linguistics. Berlin. pp. 994-1003.
[13] Qian Q, Huang M, Zhao H, Xu J, Zhu X. 2017 Assigning personality/identity to a chatting machine for coherent conversation generation. https://arxiv.org/abs/1706.02861.
[14] Bahdanau D, Cho K, Bengio Y. 2014 Neural machine translation by jointly learning to align and translate. https://arxiv.org/abs/1409.0473.
[15] Sutskever I, Vinyals O, Le QV. 2014 Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. Montreal. pp. 3104-3112.
[16] Hochreiter S, Schmidhuber J. 1997 Long short-term memory. Neural computation. 8:1735-1780.
[17] Li S, Zhao Z, Hu R, Li W, Liu T, Du X. 2018 Analogical Reasoning on Chinese Morphological and Semantic Relations. In: the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne. pp. 138-143.

[18] Ramos J. 2003 Using tf-idf to determine word relevance in document queries. In: the first instructional conference on machine learning. Piscataway. pp. 133-142.

[19] Li J, Galley M, Brockett C, Gao J, Dolan B. 2016 A Diversity-Promoting Objective Function for Neural Conversation Models. In: The North American Chapter of the Association for Computational Linguistics. San Diego. pp. 110-119.