Text Generation in Dialogue System via Deep Neural Network

Xiaodong Shi¹, Yicheng Sun¹, Yang Ding¹, Dong Han¹ and Jingyang Li²,*

¹The 28th Research Institute of CETC, Nanjing 210007, China
²School of Computer Science and Technology, Xidian University, Xi’an 710071, China

*Corresponding author e-mail: jylee@stu.xidian.edu.cn

Abstract. At present, the research on text generation focuses on the article, and there is still a large gap in the study of dialogue. At present, the task of generating sentence level has achieved good results. How to generate a complete dialogue is still challenging, and the semantic coherence involved in the dialogue is a very important issue. In view of the above problems, our research is to generate dialogue text according to dialogue elements, and construct a dialogue text generation model to realize the organization of known discrete questions and key words of answer elements into a smooth dialogue. We use the deep learning method to divide the whole dialogue text generation model into two parts. Firstly, the dialogue elements are organized into the correct dialogue process and process, and then the answer elements in the answer sentence are improved into fluent sentences. Finally, by connecting the two models, we can generate fluent multi-round dialogue texts according to the dialogue elements. We test and verify the two models, and the results show that this method can organize the dialogue process, improve the dialogue answers and generate the dialogue text.

Keywords: Dialogue Elements, Dialogue Generation, Deep Neural Network, Text Generation

1. Introduction

The generation of natural language texts similar to human expression through computer is a very important technology in the field of natural language. The generation of text level has brought a lot of research, such as the generation of article abstracts and the generation of poetry based on keywords. In the same way, multi-round dialogue has important research significance and application value. However, there are still few studies on the generation of multiple rounds of dialogue text. The purpose of dialogue text generation is to generate a smooth dialogue.

Therefore, our goal is to generate the dialogue text according to the dialogue elements, to generate a smooth dialogue according to the known discrete question set and answer element set, and to organize multiple discrete questions and answers into fluent and coherent dialogues. By analyzing the interaction mode between different roles in the dialogue, we can learn the process and rules of the whole dialogue, and solve the semantic coherence, logical coherence and sentence fluency between each question and answer sentence in the dialogue.
First of all, we select the appropriate dataset to complete the data cleaning work. Then, using the structure based on encoder decoder and using deep learning algorithm, the problem of generating dialogue text according to dialogue elements is divided into two steps. Firstly, the known discrete questions and answer elements are organized correctly, and then the key words of answer elements are expanded into complete and fluent answers, so as to organize the discrete questions and answers into coherent ones fluent dialogue text. Finally, the effectiveness of the model is tested and evaluated, and the effectiveness of the model is verified on multiple datasets.

2. Related Work
As for the organization of dialogue process, it is essentially a study of semantic coherence. At first, there was a coherence modeling method based on entity grid [1]. In recent years, inspired by deep learning, Nguyen [2] and others use convolutional neural network to extract features from solid meshes. Chen et al.[3] Proposed a sentence pair ranking model based on deep learning, modeling the binary relationship between two sentences, and using the relevant sorting algorithm to get all the order; Xu et al.[4] adding negative sampling to the sorting of sentence pairs. At the same time, some people integrated the entity information into the sentence pair ranking model [5] [6]. Li et al.[7][8] Extended the binary relationship of sentence pairs to multiple relations. Gong et al. [9] and Logeswaran et al.[10] Solved this problem based on the hierarchical cyclic network encoder decoder model. Cui et al.[11] Optimized the attention mechanism of the model.

For the task of answer generation, data driven dialogue generation [12] has become a research hotspot. Vinyals et al.[13] applied the sequence to sequence model for dialogue generation. Shang et al.[14] Introduced attention mechanism into sequence to sequence model to generate dialogue answers. However, the standard sequence to sequence model can generate a large number of responses without information. Ghazvininejad et al.[15] have encoded the external unstructured knowledge to guide the generation of response sentences. Liu et al.[16] and Zhu et al.[17] used knowledge map to integrate external structured triple information into answer sentence generation. Xing et al.[18] Introduced the topic keyword information in the answer generation of the dialogue system to make the generated answer content around the relevant topic and improve the richness of the answer content.

3. The Proposed Method
The two steps of dialogue text generation model are dialogue process model and answer generation model. We use self-attention mechanism based coding and decoding framework to realize dialogue process model, and multi-layer attention mechanism coding and decoding framework to achieve answer generation model.

3.1 Answer Generation Model
The answer generation model uses the encoder decoder framework. The overall model is divided into two modules, as shown in Fig. 1. The encoder is responsible for encoding the question and answer elements; the decoder generates the answer according to the question attention mechanism and keyword attention mechanism.
3.1.1 Encoder. The encoder encodes a long sequence using a gated loop network. First, the input question is segmented, and then each word in the input sentence and the key words of the answer elements are coded to obtain the distributed word vector of each word.

\[ \mathbf{x}_i = \mathbf{W}_w \mathbf{w}_i \]  

where \( \mathbf{w}_i \) is the one-hot coding vector of the \( i \)-th word, \( \mathbf{W}_w \) is the constructed word embedding matrix, each column represents the word vector of a word in the thesaurus, and \( \mathbf{x}_i \) is the distributed word vector of the \( i \)-th word in the sentence.

The word vector of each word in the sentence is sent to encoder GRU for encoding:

\[ \mathbf{h}_t = \text{GRU}(\mathbf{h}_{t-1}, \mathbf{x}_t) \]  

where \( \mathbf{x}_t \) is the distributed word vector of the \( t \)-th word in the sentence, \( \mathbf{h}_{t-1} \) is the hidden layer vector of the previous neuron, and \( \mathbf{h}_t \) is the output hidden layer vector of the current neuron.

3.1.2 Decoder. The decoder also uses the gated loop network as the decoder. At each step, each word is gradually generated according to the decoder's state, dialogue questions and key words of target answer elements.

Based on the multi-layer attention mechanism, the decoder needs the current hidden state \( \mathbf{h}_t \), question vector \( \mathbf{c}_t \) and keyword vector \( \mathbf{k}_t \) when generating each word:

\[ \mathbf{h}_t = \text{GRU}(\mathbf{h}_{t-1}, \mathbf{W}_e \mathbf{o}_{t-1}) \]  

\[ \mathbf{a}_t = \text{Linear}(\mathbf{h}_t; \mathbf{c}_t; \mathbf{k}_t) \]  

\[ \mathbf{p}_t = \text{softmax}(\mathbf{a}_t) \]  

\[ \mathbf{o}_t = \text{argmax}(\mathbf{p}_t) \]  

\[ \text{Linear}(\mathbf{x}) = \mathbf{W}_x \mathbf{x} + \mathbf{b} \]  

where \( \mathbf{o}_{t-1} \) is the word selected in the last time step and expressed in the form of a one-hot vector. \( \mathbf{W}_e \) is the constructed word embedding matrix, \( \mathbf{h}_{t-1} \) is the output hidden layer vector of the previous neuron, and \( \mathbf{h}_t \) is the hidden layer vector of the current neuron output. \([\mathbf{h}_t; \mathbf{c}_t; \mathbf{k}_t]\) is the operation of
splicing the three vectors, and \( \mathbf{a}_i \) is the vector obtained by mapping the vector to the vocabulary dimension. \( \mathbf{o}_i \) is the selected word with the maximum probability, and the one-hot vector is used to indicate the serial number of the word in the vocabulary.

The question vector \( \mathbf{c}_i \) represents the semantic information, which is obtained through the attention mechanism of the question:

\[
\beta_i = \mathbf{h}_i^T \mathbf{W}_c \mathbf{s}_i \\
\omega = \text{softmax}(\mathbf{\beta}) \\
\mathbf{c}_i = \sum_i \omega_i \mathbf{s}_i
\]

where Formula (22) is the similarity calculation formula, \( \mathbf{W}_c \) is the training parameter of the model, and \( \beta_i \) is the correlation score of each vector output by the encoder. \( \mathbf{\beta} \) is the vector of each \( \beta_i \), and \( \mathbf{\omega} \) is the weight vector after normalization. \( \mathbf{s}_i \) is the \( i \)-th output vector of the encoder, and \( \mathbf{c}_i \) is the question vector under the attention mechanism in the \( i \)-th decoder step.

The keyword vector represents the semantic information provided by \( \mathbf{k}_i \), representing the answer elements:

\[
\beta_i = [\mathbf{s}_i; \mathbf{h}_i]^T \mathbf{W}_k \mathbf{e}_i \\
\omega = \text{softmax}(\mathbf{\beta}) \\
\mathbf{k}_i = \sum_i \omega_i \mathbf{e}_i
\]

where Formula (25) is the similarity calculation formula, \( \mathbf{W}_k \) is the training parameter of the model, and \( \beta_i \) is the correlation score of each answer element vector.

4. Experiments and Evaluation

The quality of the generated dialogue text depends on the correctness of the model organization process and the quality of the generated answers. Therefore, we evaluate the model according to the correctness of the dialogue process and the quality of the answers generated by the answer elements.

4.1 Dialogue Process Organization

4.1.1 Dataset. For the test of dialogue process organization, we use three datasets, and now we introduce each dataset.

1) The E-commerce-1 dataset is obtained from the Chinese E-commerce dataset [19]. Keywords are extracted from the answers of 200 000 groups of conversations in the E-commerce dataset to construct a new 200 000 groups of dialogue data. After mixing the two parts, 300 000 groups are randomly selected as the training set, and 1 000 groups of conversations are randomly extracted from the remaining part as the test set.

2) The Accident dataset, which is an English news accident report chapter, contains only 200 chapters. Therefore, 180 groups are extracted as training set and 20 groups as test set. This dataset only tests the PM value of sentence contrast rate with correct relative position. The test set under this index contains about 1000 sentence pairs, which can provide this metric.

3) NIPS Abstract dataset, which is a summary of 3 272 documents extracted from nips, uses 3 000 passages as training set and 272 passages as test set.
For the above three datasets, the word vector of E-commerce-1 dataset is trained in the process of training; for the Accident and NIPS Abstract dataset, the pretrained 300 dimensional Glove word vector is used. And we extract specific part of speech words as keywords of answer elements.

**Table 1.** Statistics of three datasets used in our experiments

| Dataset          | Train | Test |
|------------------|-------|------|
| E-commerce-1     | 300,000 | 1,000 |
| Accident         | 180    | 20   |
| NIPS Abstract    | 3,000  | 272  |

4.1.2 Result. The word vector dimension and the hidden layer vector dimension of the dialogue process model are set to 300, the attention mechanism layer is set to 3 layers, the FNN layer dimension is 600, and the initial learning rate is 0.001, which gradually decreases with the training process. For the accident dataset, the batch size is 8 and the thesaurus size is 3892; for the nips Abstract dataset, the batch size is set to 16 and the thesaurus size is 13128; for the e-commerce-1 dataset, the batch size is set to 128, and the thesaurus size is 17100.

We compare the model with other related models, and the test results are shown in Table 2. Compared with the rule-based Entity Grid Model and the deep learning based Seq2seq model, the effect of the model is significantly improved. However, compared with the same generative model LSTM with PtrNet model, the effect in the Accident dataset is better than that of the model, but the effect is not good in the NIPS Abstract dataset.

**Table 2.** Results of dialogue process organization

| Model            | Accident | NIPS Abstract | E-commerce-1 |
|------------------|----------|---------------|--------------|
|                  | PM       | Acc           | PMR          | Acc | PM |
| Entity Grid      | 0.56     | 0.201         | 0.54         | -   | -  |
| Seq2seq          | -        | 0.272         | 0.63         | -   | -  |
| LSTM+PtrNet      | 0.79     | 0.508         | 0.83         | -   | -  |
| Our Model        | 0.81     | 0.478         | 0.82         | 0.265 | 0.512 | 0.63 |

4.2 Answer Generation

4.2.1 Dataset. The E-commerce-2 dataset is extracted from E-commerce for the task of generating answers. The keywords extracted from the dialogue in E-commerce are taken as the answer elements, and a ternary dataset of questions, answer elements and standard answers is constructed. 200000 sets are extracted as training set and 1200 groups are used as test set.

4.2.2 Metric. There are two indicators used in the quality test of answer generation.

1) BLEU: BLEU [20] indicator is used to evaluate the difference between the sentences generated by the model and the actual sentences. For each phrase in the generated sentence, the ratio of the simultaneous occurrence in the actual sentence is calculated, which is the BLEU score. According to the number of words in a phrase, Bleu can be calculated from different scales.

2) ROUGH: this indicator is similar to BLEU, which calculates the recall rate and the ratio of phrases in actual sentences to the generated sentences.

4.2.3 Result. The initial learning rate was 0.001, and the size of vocabulary was 19000. The E-commerce-2 dataset was iterated for 35 rounds, and the batch size was set to 128.

The evaluation of answer generation quality is divided into automatic test and manual test. The BLEU and ROUGH values are calculated by automatic test, and we compared our model with S2SA, the sequence to sequence model with attention mechanism, and the manual test is used to judge the quality of answer generation artificially. The results of the automatic test are shown in Table 3.
Table 3. Automatic evaluation of answer generation quality

| Model  | BLEU-1 | BLEU-2 | BLEU  | ROUGH-1 | ROUGH-2 | ROUGH |
|--------|--------|--------|-------|---------|---------|-------|
| S2SA   | 17.61  | 9.64   | 0.41  | 16.93   | 9.89    | 0.40  |
| Our Model | 33.86  | 15.55  | 2.32  | 30.51   | 15.74   | 2.35  |

Table 4. Human evaluation of answer generation quality

| Score | Percentage |
|-------|------------|
| 2     | 52%        |
| 1     | 23%        |
| 0     | 25%        |

5. Conclusion
In this task, we first determine the dialogue text generation task, and design a feasible research scheme for the research purpose. The overall purpose is divided into two parts: the dialogue process model and the answer generation model. A dialogue process model based on self-attention mechanism is proposed, which can organize the dialogue elements into a reasonable dialogue process; and a multi-layer attention mechanism based answer generation model is proposed, which can improve the answer elements into smooth answers. The two models solve the dialogue process and answer generation respectively, so as to achieve the final purpose of generating dialogue. The experimental results show that our model can effectively generate fluent dialogue text according to the dialogue elements.

References
[1] Barzilay R., Lapata M. Modeling local coherence: An entity-based approach. Computational Linguistics, 2008, 34(1): 1-34.
[2] Nguyen T., Joty S. A neural local coherence model. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 1320-1330.
[3] Chen X., Qiu X., Huang X. Neural sentence ordering. arXiv preprint arXiv:1607.06952, 2016.
[4] Xu P., Saghir H., Kang S. A Cross-Domain Transferable Neural Coherence Model. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 678-687.
[5] Du S., Xu F., Wang Mi. An Entity-driven Bi-directional LSTM Model for Discourse Coherence. Journal of Chinese Information Processing, 2017, 031(6):67-74.
[6] Luan K., Du X., Sun C., Liu B., Wang X. Sentence Ordering based on Attention Mechanism. Journal of Chinese Information Processing, 2018, 32(01):123-130.
[7] Li J., Hovy E. A model of coherence based on distributed sentence representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014: 2039-2048.
[8] Li J., Jurafsky D. Neural Net Models of Open-domain Discourse Coherence. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 198-209.
[9] Gong J., Chen X., Qiu X. End-to-end neural sentence ordering using pointer network. arXiv preprint arXiv:1611.04953, 2016.
[10] Logeswaran L., Lee H., Radev D. Sentence ordering and coherence modeling using recurrent neural networks. Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
[11] Cui B., Li Y., Chen M. Deep attentive sentence ordering network. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 4340-4349.
[12] Chen H., Liu X., Yin D. A survey on dialogue systems: Recent advances and new frontiers. Proceedings of ACM SIGKDD, 2017, 19(2): 25-35.
[13] Vinyals O., Le Q. A neural conversational model. arXiv preprint arXiv:1506.05869, 2015.
[14] Shang L., Lu Z., Li H. Neural Responding Machine for Short-Text Conversation. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015: 1577-1586.
[15] Ghazvininejad M., Brockett C., Chang W. A knowledge-grounded neural conversation model. Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
[16] Liu S., Chen H., Ren Z. Knowledge diffusion for neural dialogue generation. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 1489-1498.
[17] Zhu W., Mo K, Zhang Y. Flexible end-to-end dialogue system for knowledge grounded conversation. arXiv preprint arXiv:1709.04264, 2017.
[18] Xing C., Wu W., Wu Y. Topic aware neural response generation. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017: 3351-3357.
[19] Zhang Z., Li J., Zhu P. Modeling Multi-turn Conversation with Deep Utterance Aggregation. Proceedings of the 27th International Conference on Computational Linguistics. 2018: 3740-3752.
[20] Papineni K., Roukos S., Ward T. BLEU: a method for automatic evaluation of machine translation. Proceedings of the 40th annual meeting on association for computational linguistics, 2002: 311-318.