

Using fiction to enhance topic-specific conversation generation

Jiancheng Lin Chao Huang Gang Qiao Xinyi Liu Paiming Liu Zhenpeng Zhan

GBU, BaiduInc, China, 518052

{linjiancheng01, huangchao17, qiaogang01, liuxinyi07, v_liupaiming, zhangzhenpeng01}@baidu.com

1 Abstract

In recent years, open-domain dialogue systems have made remarkable achievements in the large-scale pre-trained dialogue models. Nevertheless, constructing a topic-specific dialogue system continues to confront the challenge of acquiring high-quality dialogue data, which is inherently difficult to collect. We propose a method for improve dialogue generation quality using fiction to address this problem. Our method uses fiction on relevant topics, converts the text data into single-party dialogue, and finetunes the model after mixing it with a multi-party dialogue corpus. In our paper, we focus on the topic of gender relations, which is difficult to get relevant datasets, for the experiment. According to our experiments, on the basis of the original model capability, our method not only effectively improves trigger rate of the model in a specific topic, but also increases the model response information quantity.

2 Introduction

The field of dialogue systems has witnessed significant advancements in recent years, especially with the emergence of large-scale pre-trained models. These models demonstrate remarkable capabilities in generating human-like responses and engaging in meaningful conversations. The progress in natural language understanding and generation has propelled the development of dialogue systems and has opened up new opportunity for various applications, such as customer service, virtual assistant, and social chatbot.

Prominent examples include GPT-3, Plato series, and Meena, which have showcased the potential in generating coherent and contextually appropriate responses. However, despite the remarkable progress, dialogue systems still face challenges in constructing topic-oriented conversations. They often lack a deep understanding of specific domains or topics. Constructing meaningful dialogues that are centered around a particular subject requires access to high-

quality, topic-specific data with rich and informative content. Unfortunately, such data is often scarce or unavailable, limiting the performance and practicality of dialogue systems in specialized domains.

In this paper, we propose a fiction approach to address this challenge by leveraging fiction to enhance the effectiveness of topic-specific dialogue generation. Our method transforms fiction text into single-party dialogues. Because of the abundant details and scene descriptions in the fiction, converted single-party dialogues can contain greater information quantity on a specific topic. Based on our experimental results, our approach effectively enhances the informativeness of topic-oriented dialogues and increases the trigger rate for the model to enter the desired topic.

3 Related work

In recent years, open-domain dialogue systems such as Meena, Blenderbot, and Plato have made significant progress in generating human-like responses. However, there are several problems to be solved in the practical application of dialogue systems. Among them, the lack of knowledge and informativeness of dialogues on specific topics is the main problem.

To address this issue, past research has focused on pre-training phase scaling model and dataset sizes to achieve better performance. For example, PaLM with 540B parameters and GPT-3 with 173B parameters have demonstrated powerful zero/few-shot knowledge in quizzing capabilities. However, in practical applications, it is difficult to deploy these models. During fine-tuning, catastrophic forgetting may occur, and there is a lack of high-quality, information-rich corpus on the Internet about specific topics, such as gender relations. For this reason, our paper proposes the use of fictional data in the fine-tuning phase to enhance dialogue generation.

4 Methodology

The following section describes the construction method proposed in our paper. The proposed approach uses topic-specific web texts to enhance the dialogue generation on a specific topic. Web texts including news, fictions, encyclopaedias are richer and more in-depth than dialogue corpora. The use of a topic-specific generic corpus can provide in-depth content and rich information to the conversations on that topic. We constructed a topic-specific single-party conversational corpus using fiction data related to the topic of the conversation.

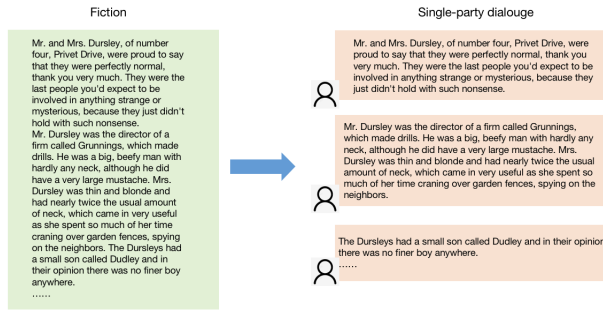


Figure 1: Transforming a fiction into a single party dialogue

As figure 1 shows, there are short stories from the Internet with details of action description and scene description on specific topic.

The fiction data was first divided into sentences using symbols such as newline symbol. After that, we set the minimum length of N , and multiple words less than N are combined into a single sentence, and the sentence whose words number is larger than N is retained. Multiple sentences are constructed together into a dialogue session, and each sentence is regarded as an utterance in the session.

We use the same training method as the work [1]. Specifically, We mix single-party and multi-party dialogue and then finetune the Prefix LM. The model input includes four parts: token embedding, type embedding, role embedding, and position embedding. Different role embedding is used according to the speaker in multi-party dialogue, and the same role embedding is used in single-party dialogue.

Unlike the work [1], which uses Web Text as single-party dialogue in a pre-trained phase to learn knowledge, this paper learns the ability to describe details in the fiction data during the fine-tuning phase to enhance the informativeness of the model responses.

5 Experiments

5.1 Settings

In the experiments in this paper, we do not use topics such as film, music as there is enough publicly available data on these topics. We use gender relations as a dialogue topic to demonstrate the validity of our proposed approach. We collected 8,999 fictions from the Internet, including romance fictions and sex-related fictions.

For our experiments, we selected plato-k [1] as the base model. Plato-k is a model trained using both dialogue corpus and web text, which reaches SOTA in the dialogue area. In the pre-training of plato-k, our multi-party dialogue datasets consisted of japanese dialogue datasets that are publicly available on social media. Web texts data are used from Wiki and Oscar japanese version datasets. We pre-train two plato-k models of different sizes, plato-k (24L) and plato-k (32L), respectively. The structure and parameters of the models are shown in the Table 2.

Besides, we also use plato-2 [2] as a baseline model to illustrate the effectiveness of the method proposed in this paper. Plato-2 has common parameters with plato-k in table 2, except that the position embedding used is different from that of plato-k. Plato-2 uses pure conversational corpus for training and reaches SOTA in Chinese and English. We found that the pre-trained models are easy to produce simple, uninformative responses on the topic of gender relations, or do not follow the user's intention. Therefore, in our experiments, the following metrics were used to manually and automatically evaluate the responses of different models to determine the effectiveness and generalizability of the approach proposed in this paper:

1) Length of model responses. We automatically count the response lengths generated by the models fine-tuned using our method as a metric to evaluate the amount of information in the model responses.

2) Trigger rate. We collected 200 volunteers' chat data on the topic of gender relations, including 100 multi-round data and 100 single-round data. The data were used as contextual input to the model to eval if the model followed.

3) Coherence and knowledge quiz. We did not want our finetune model to lose its original capabilities of the model, so we still manually conducted chitchats and knowledge quizzes with the model to evaluate the basic capabilities. Coherence is evaluated at three levels: 0, 1, and 2, where 0 means incoherent and 2 means fully coherent. Knowledge quizzes also is evaluated at three levels: 0, 1, and 2, where 0 means less than 20% of the general knowledge quizzes are accurate, 1 means 50% of the general knowledge quizzes are accurate and 2 means more than 50% are accurate.

Table 1: Model evaluation

model	tigger rate	response length	coherence	knowledge
plato-2(24L)	64.65%	14.99	2	0
plato-k(24L)-pure-dialogue	55.8%	13.06	2	1
plato-k(24L)	65.4%	15.40	2	1
plato-k(32L)	64.7	16.92	2	1

Table 2: Model parameters

	plato-k(24L)	plato-k(32L)
layer	24	32
hidden size	1024	2048
attention heads	16	32
vocab size	32000	32000

Table 3: Pretrained Model Evaluation

	plato-2	plato-k
average response length	13.2	14.6
Knowledge quiz accuracy	6%	26.3%
Topic trigger rate	60.3%	59.8%

5.2 Experimental Results

As shown in Table 3, in the comparison of the pre-trained models, the knowledge accuracy of plato-k is 20.3% higher than that of plato-2, which is trained with only conversational corpus, and the average word length is improved by 1.4 words because of the use of generic corpus. The topic triggering rate, however, is slightly lower than that of plato-2.

We used the method proposed in our paper for fine-tuning in plato-2 and plato-k. As a comparison of the results, we used another plato-k model for fine-tuning pure multi-party dialogues, denoted as plato-k-pure-dialogue. From the table 1, we know that using pure multi-party dialogue fine-tuning brings no benefit. This may be due to the fact that our corpus of topic-special dialogue is selected from social media using keywords with limited coverage and that these conversations may have already appeared during pre-training.

However, compared to fine-tuning using pure multi-party dialogue data, fine-tuning using a mixture of single-party dialogue and multi-party conversation data improves the trigger rate by 9.6% and the average response length of sentence responses by 2.34.

Using different base models the use of fiction to construct single-party both resulted in significant benefits. Comparing the same model in the manual evaluation of coherence and informativeness using only single-party dialogue fine-tuning and using multi-party dialogue mixed with single-party dialogue fine-tuning did not result in significant differences.

As shown in Table 1 and Table 2, pre-trained models are more likely to benefit in terms of the amount of

Table 4: The win rate of model

	win	tie	lose
plato-k(32L) vs plato-k(24L)	18.04%	69.58%	12.37%

information when a larger number of parameters are used. This suggests that models with a larger number of participants are more likely to benefit from a single-party dialogue.

6 Conclusion

In order to address the difficulties caused by the lack of high-quality dialogue corpus for building topic-specific dialogue systems, this paper proposes a method to enhance dialogue generation using fiction data. The method uses fiction data on relevant topics, which are transformed into single-party dialogues using the method proposed in this paper, and fine-tuned in the model together with multi-party dialogues. According to our experimental performance, our method can effectively improve the triggering rate of the model on a certain topic dialogue and the amount of information the model replies. Also, the proposed method can achieve better performance on models pre-trained with multi-party dialogue mixed with single-party dialogue data compared to models pre-trained with pure multi-party dialogue. Our experiments were conducted in Japanese, and we achieved similar results in English. In the future, our work will focus on multilingual models.

References

- [1] Siqi Bao, Huang He, Jun Xu, Hua Lu, Fan Wang, Hua Wu, Han Zhou, Wenquan Wu, Zheng-Yu Niu, and Haifeng Wang. Plato-k: Internal and external knowledge enhanced dialogue generation. *arXiv preprint arXiv:2211.00910*, 2022.
- [2] Siqi Bao, Huang He, Fan Wang, Hua Wu, Haifeng Wang, Wenquan Wu, Zhen Guo, Zhibin Liu, and Xinchao Xu. Plato-2: Towards building an open-domain chatbot via curriculum learning. *arXiv preprint arXiv:2006.16779*, 2020.