Exploring Query Reformulation for Conversational Information Seeking

Disen Wang, Hui Fang,

University of Delaware, Newark DE 19716, USA

Abstract: Few tasks have been designed for conversational information seeking. To fill this gap, this year's TREC Conversation Assistance Track (CAsT) is proposed to advance research on conversational search systems. We built a model that first read the dialogue context, then retrieved candidate response information from a large collection of paragraphs. In order to perform passage retrieval task, we first applied the coreference resolution method to format questions into queries, and we use Indri to retrieve top 100 relevant passages. During the second phrase, we applied fine-tuned BERT model to rerank retrieved passage.

1 Introduction

The TREC conversational assistance track (CAsT)[1] is introduced with the primary initial focus on understanding information needs in a conversational format and finding relevant responses with contextual information. This year the track focus on retrieval of relevant result content in context. Specifically, the task is retrieval-based candidate response ranking in context. Given the current dialogue turns up to the given turn, the model is asked to retrieve candidate text pas-

[1] http://www.treccast.ai

```
Title: forms of energy
Description: Energy and its various forms.
1       What are the different forms of energy?
2       How can it be stored?
3       What type of energy is used in motion?
4       Tell me about mechanical energy.
5       Give me some examples.
6       Why is sound a form of mechanical energy?
7       How does it differ from potential energy?
8       Are potential and kinetic the same?
9       What type does chemical energy belong to?
10      What form of energy is used in eating?
```

Figure 1: Example of questions under topic ”forms of energy”.

sages from a fixed text collection for the current turn. Figure 1 shows example of sequentially-related turns under certain topic.

Based on the task, we proposed a method that consists of two key steps: query formatting and passage re-ranking. More specifically, we first applied the coreference resolution to formulate each turn into a query, and then we use Indri search engine to retrieve top 100 relevant passages. After that, we applied fine-tuned BERT (*2*) model to re-rank retrieved passages and choose top 30 passages as outputs.

2 Method

The corpus is a combination of three standard TREC collections: MARCO Ranking passages (*1*), Wikipedia (TREC CAR) (*3*), and News (Washington Post). Because the corpus contains a large amount of passages, we chose to use two-phase retrieval: first retrieval top 100 relevant passages with Indri, and then re-rank retrieved passages with fine-tuned BERT model. The key step in the first phase is to format query, and the key step in the second step is to fine-tune the BERT model based on labeled dataset.

2.1 Query Formatting

The original questions under each topic are written in natural language. In order to retrieve relevant passages we need to format each question into a query. It is worth noting that the questions are sequentially related, several consecutive questions might be asking about same item, and later questions will tend to use pronouns to refer to the nouns shown in previous question. In order to generate queries with high quality, it is essential to replace those pronouns with the nouns they are referring to. We applied the NeuralCoref model[2] to compute the probability between each pronoun in current question and each nouns in previous question, and replace the pronouns with the corresponding nouns with highest probability. After coreference resolution step, we will remove stop words to form a temporary query. Please note that for each question, we can only leverage information from previous questions. Therefore, we can not use any information from later questions.

Besides, we noticed that for each topic, the organizers also provide the title and description. We believe the title and description could help us to narrow down the area of current question. However, the description is too specific that it can be misleading if we directly extend each query with description. Dut to the time constraint, we decided to extend each query with topic title to improve the information retrieval performance, especially when original query contains few words. Specifically, when formatting the final query, we assigned the words from original query with weight equals to 0.8 and the words from title with weight equals to 0.2. After

[2]https://github.com/huggingface/neuralcoref

formatting queries, we retrieved top 100 relevant passages with Indri (*4*), and re-rank retrieved passages in next phrase.

2.2 Fine-tuning BERT Model

BERT model is the state of the art language model for many NLP tasks, including MSMARCO passage re-ranking task [3]which is very similar to this task. Besides, it is powerful when adapting to new tasks by fine-tuning parameters. Therefore, we choose to fine-tune the BERT model with labeled data and use fine-tuned BERT model to re-rank retrieved passages in first phrase. The organizers provide labels for 5 topics (approximately 50 question and 500 passages in total). The judgements are graded on a three point scale (2 very relevant, 1 relevant and 0 not relevant). We first regularize the points into 1 scale: 1 very relevant, 0.5 relevant and 0 not relevant, then we fine-tuned the pre-trained BERT model based on training labels. After getting fine-tuned model, we applied BERT model to re-rank retrieved passages.

3 Submitted Runs

We submitted 2 runs for this year's track. The details of each submitted run is introduced as below:

UDInfoC_BL: The baseline run where we applied the NeuralCoref model for coreference resolution and use Indri to retireve relevant passages.

UDInfoC_TS: The run where we applied the NeuralCoref model for coreference resolution and use Indri to retireve relevant passages, then we re-ranked the retrieved passages witn fine-tuned BERT model. Because running BERT requires huge amount of compute resource, we only re-ranked passages from first 30 topics in test set. Therefore, in the final submitted run only first 30 topics were re-ranked, and other topics are the ranking results from Indri.

[3]http://www.msmarco.org

	MAP@5	DNCG@5
Median System	0.042	0.296
Baseline Method	0.075	0.311
Transfer Learning	0.061	0.259

Table 1: Performance Comparison

4 Evaluation

Table 1 shows the performance of the baseline method, transfer learning method and the median results for the mean across all turns assessed. As can be seen from Table 1, baseline method performs better than the median system on both MAP and DNCG evaluation metrics, which illustrates the effectiveness of the query formatting strategy we proposed. However, the transfer learning method performs worse than the baseline method, which may be the result of limited amount of available labeled training data.

5 Conclusion

In this year's conversational assistant track, we first applied coreference resolution method to format questions into queries, then we use Indri to retrieve top 100 relevant passages. During the second phrase, we use fine-tuned BERT model to re-rank retrieved passage. Evaluation results show that the baseline method performs better than the median results of the challenge, which illustrates the effectiveness of our query format strategy. On the other hand, with limited labeled training data available, the performance of transfer learning method is worse than the baseline method. We anticipate that with more training data available the transfer learning method would yield to better performance.

References and Notes

1. Bajaj, P., Campos, D., Craswell, N., Deng, L., Gao, J., Liu, X., Majumder, R., McNamara, A., Mitra, B., Nguyen, T., Rosenberg, M., Song, X., Stoica, A., Tiwary, S., and Wang, T. Ms marco: A human generated machine reading comprehension dataset, 2016.

2. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

3. Dietz, L., Verma, M., Radlinski, F., and Craswell, N. Trec complex answer retrieval overview. In *TREC* (2017).

4. Strohman, T., Metzler, D., Turtle, H., and Croft, W. B. Indri: A language model-based search engine for complex queries. In *Proceedings of the International Conference on Intelligent Analysis* (2005), vol. 2, Citeseer, pp. 2–6.