BECON: BERT with Evidence from CONceptNet for Common Sense Question Answering

Anonymous ACL submission

Abstract

CommonsenseQA (Talmor et al., 2018) is created by crowdsourcing workers based on knowledge graphs on ConceptNet (Speer et al., 2017). Solving the task requires the model to have common sense or world knowledge like humans. Current LM-pretrained model such as BERT (Devlin et al., 2018) achieves state-of-the-art performance on the CQA dataset, which implies that language models trained on very large corpus may learn some sort of common sense knowledge implicitly. On the other hand, with the availability of the large knowledge graph such as ConceptNet, which contains explicit common sense knowledge, we would like to investigate how to use the explicit form of common sense knowledge as complementary to BERT which has implicit common sense knowledge.

For CommonsenseQA task, a question and five candidate answers are given, and one of the five answers is correct. The candidate asnwers usually consist of one or two words, forming a *concept*. According to (Talmor et al., 2018), the best performing baseline model is the BERT-large model finetuned with CQA dataset. Our model is built based on BERT-large model, but also utilizes additional evidences from ConceptNet, which provides useful information to answer the question.

Concretely, to use the knowledge in Conceptnet, we first query each candidate answer in ConceptNet to get a list of evidence sentences which may be helpful to answer the question. In order to reduce the noise, we use pretrained BERT with Next-Sentence-Prediction head (BERT-NSP) to rank the evidence sentences and select the top-scored one. A example of the evidence sentences are shown in Figure 1. We believe that BERT-NSP is helpful to rank the relevancy of the question and the evidence sentence. As in (Talmor et al.,

What is a wet person likely to do?	
A. suicide	Jumping out of a window is for suicide
B. catch cold	*Something that might happen as a consequence of getting wet is you catch a cold*
C. cross street	One of the things you do when you cross the street is look both ways
D. gain weight	an overeating individual can gain weight
E. thank god	a person can thank God

Figure 1: A question from CommonsenseQA dataset and its 5 candidate answers with their corresponding top-ranked evidences. The correct answer are in **bold**/green, and the evidence corresponding to the correct answer is it *italic*.

Model	test F1
BERT-large (Talmor et al., 2018)	56.7
CoS-E (Rajani et al., 2019)	58.2
BECON	57.9
BECON (ensemble)	59.6

Table 1: Comparison of the test accurary with literature.

2018), each question-answer pair is linearized into a delimiter-separated sequence (i.e., "[CLS] If ... ? [SEP] bedroom [SEP]") and the hidden vector over the [CLS] token are used as representation of the choice. For our BECON model, we further concatenate the evidence sentence (i.e., "[CLS] If ... ? [SEP] bedroom [SEP] bedroom is a place for sleeping [SEP]"), which may help the model make better decisions.

The experiment results on CQA test split are shown in Table 1. Our single model outperforms the BERT-large baseline by 1.2%. Using ensemble technique, our model achieves 59.7%, outperforms CoS-E (Rajani et al., 2019) by 1.4%.

In conclusion, our model shows that explicit common sense knowledge can be complementary to BERT in solving common sense question answering tasks.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Nazneen Fatema Rajani, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Explain yourself! leveraging language models for commonsense reasoning. *arXiv preprint arXiv:1906.02361*.

Robert Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. 2018. Commonsenseqa: A question answering challenge targeting commonsense knowledge. *arXiv preprint arXiv:1811.00937*.