IIE-NLP-Eyas at SemEval-2021 Task 4: Enhancing PLM for ReCAM with Special Tokens, Re-Ranking, Siamese Encoders and Back Translation

Yuqiang Xie, Luxi Xing, Wei Peng and Yue Hu[§]

Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

Introduction

This paper describes approaches for all three subtasks of SemEval-2021 Task 4: Reading Comprehension of Abstract Meaning, which developed by the IIE-NLP-Eyas Team (Natural Language Processing group of Institute of Information Engineering of the Chinese Academy of Sciences). Our contributions are summarized as the followings:

1. We design many simple and effective approaches to improve the performance of the PLMs on all three subtasks, such as adding special tokens, sentence re-ranking and so on;

2. Experiments demonstrate that the proposed methods achieve significant improvement compared with the PLMs baseline and we obtain eighth rank (**87.51%**) and tenth rank (**89.64%**) on the official blind test set of subtask 1 and subtask 2 respectively.

Task Definition

Formally, suppose there are seven key elements in all subtasks, i.e. $\{D, Q, A1, A2, A3, A4, A5\}$.

1. D denotes the given article.
2. Q denotes the summary of the article with a *placeholder*.
3. A_* the candidate abstract concepts for all sub- tasks to fill in the *placeholder*.

Methods

1. Multi-Choice Based Model
2. Special Tokens
3. Sentence Ranking
4. Siamese Encoders
5. Back Translation
6. Label Smoothing

Contact Information

- Code available at https://github.com/indexfziq/IIE-NLP-Eyas-SemEval2021
- Email: xieyuqiang@iie.ac.cn
- Phone: +86 1591 0702 975

Multi-Choice Based Model

A multiple-choice based QA model $\mathcal{M}$ consists of a PLM encoder and a task-specific classification layer which includes a feed-forward neural network $f(\cdot)$ and a softmax operation. For each pair of question-answer, the calculation of $\mathcal{M}$ is as follow:

$$score_i = \frac{exp(f(S_i))}{\Sigma_{i'} exp(f(S_{i'}))} \quad (1)$$

$$S_i = \mathrm{PLM}([Q; A_i; D]) \quad (2)$$

where the $[\cdot]$ is the input constructed according to the instruction of PLMs, and the S_* is the final hidden state of the first token (<s>).

Special Tokens

To help the model to the PLMs represent and understanding the abstract concept in textual descriptions, we use <e> and </e> to add on both ends of the abstract concept, i.e.:
`<e> abstract concept </e>`.

Sentence Ranking

To rank the sentences in D, we resort BERT to compute the similarity score between each sentence, i.e. W_n, and Q following the algorithm in bert-score. After ranking, the sentences in D are sorted in descending order of similarity scores, and we can get a rearranged passage $\hat{D}$ as the passage input to the QA model.

Siamese Encoders

We propose a siamese encoders based architecture to inject the additional complete question statement information while not influence the input with passage. On the other hand, it can be seen as introducing an auxiliary task to assist the main task.

$$l_i^1 = PLM([\hat{Q}_i])[0] \quad (3)$$

$$l_i^2 = PLM([Q; A_i; D])[0] \quad (4)$$

$$P^1(A_i|\hat{Q}) = \mathrm{softmax}(f(l_i^1)) \quad (5)$$

$$P^2(A_i|D, Q) = \mathrm{softmax}(f(l_i^2)) \quad (6)$$

Back Translation

We use the Google API[‡] to translate the passage into French, and then translate the translation into English in turn. The pseudo parallel corpus can be obtained as:

$$\{D'\} = bkt(\{D\}) \quad (7)$$

where $\{D'\}$ means the translated English corpus that we used as data augment, bkt is back translation.

Label Smoothing

To improve the generalization ability of the model trained on sole task and prevent the overconfidence of model, we consider training model with label smoothing. When training with label smoothing, the hard one-hot label distribution is replaced with a softened label distribution through a smoothing value α, which is a hyper-parameter. In our experiments, we set the smoothing value $\alpha = 0.1$.

Results

Models	Trial Acc.	Dev Acc.
RoBERTa$_{\text{LARGE}}$	85.85	82.12
(1) w/ special tokens	**87.81**	**87.69**
(2) w/ sentence ranking	86.54	83.52
(3) w/ label smoothing	86.88	85.85
(4) w/ siamese encoders	86.62	83.22
(5) w/ back translation	87.23	84.32
Our Approach	**87.81**	**87.69**

Table 1: The results of on Imperceptibility.

Models	Trial Acc.	Dev Acc.
RoBERTa$_{\text{LARGE}}$	**88.51**	85.93
(1) w/ special tokens	87.47	88.98
(2) w/ sentence ranking	87.29	86.84
(3) w/ label smoothing	87.67	87.08
(4) w/ siamese encoders	87.34	86.18
(5) w/ back translation	88.41	87.54
Our Approach	87.10	**89.54**

Table 2: The results of on Nonspecificility.

Trained on	Tested on	Test Acc.
Subtask-1	Subtask-1	87.51
Subtask-1	Subtask-2	84.13
Subtask-2	Subtask-2	89.64
Subtask-2	Subtask-1	81.09

Table 3: The results on Interaction.

Analysis

Special Token	Trial Acc.	Dev Acc.
<e> </e>	88.01	**87.10**
<#> </#>	**88.63**	86.93
<$> </$>	88.12	86.26
# /#	87.34	85.89
$ /$	87.73	86.13
N/A	86.23	83.12

Table 4: The results of models with different special tokens on Imperceptibility Task.

Important Result

Compared to the backbone model RoBERTa large model, our methods achieve significant improvements. It is interesting that the special token is the most helpful part for both Imperceptibility and Nonspecificility subtasks. Label smoothing works well for avoiding over-fitting of PLMs.

Conclusion

In this paper, we design many simple and effective approaches to improve the performance of the PLMs on all three subtasks. Experiments demonstrate that the proposed methods achieve significant improvement compared with the PLMs baseline and we obtain the eighth-place in subtask-1 and tenth-place in subtask-2 on the final official evaluation. Moreover, we show that special tokens are useful features contributing to most of the system's boost, which work well in enhancing PLMs for representating and understanding abstract concepts.

Acknowledgements

We thank the anonymous reviewers for their insightful feedback. This work has been supported by the National Key Research and Development Program of China under Grant No.Y750211101.