Building Sequential Inference Models for End-to-End Response Selection

Jia-Chen Gu[1], Zhen-Hua Ling[1], Yu-Ping Ruan[1] and Quan Liu[1,2]
[1]National Engineering Laboratory for Speech and Language Information Processing,
University of Science and Technology of China
[2]iFLYTEK Research

Introduction

Track 1 of the 7th Dialogue System Technology Challenges (DSTC7) is a kind of **retrieval-based** task which selects the correct response for the current conversation from a set of candidates with response selection algorithms. Two kinds of datasets are provided.

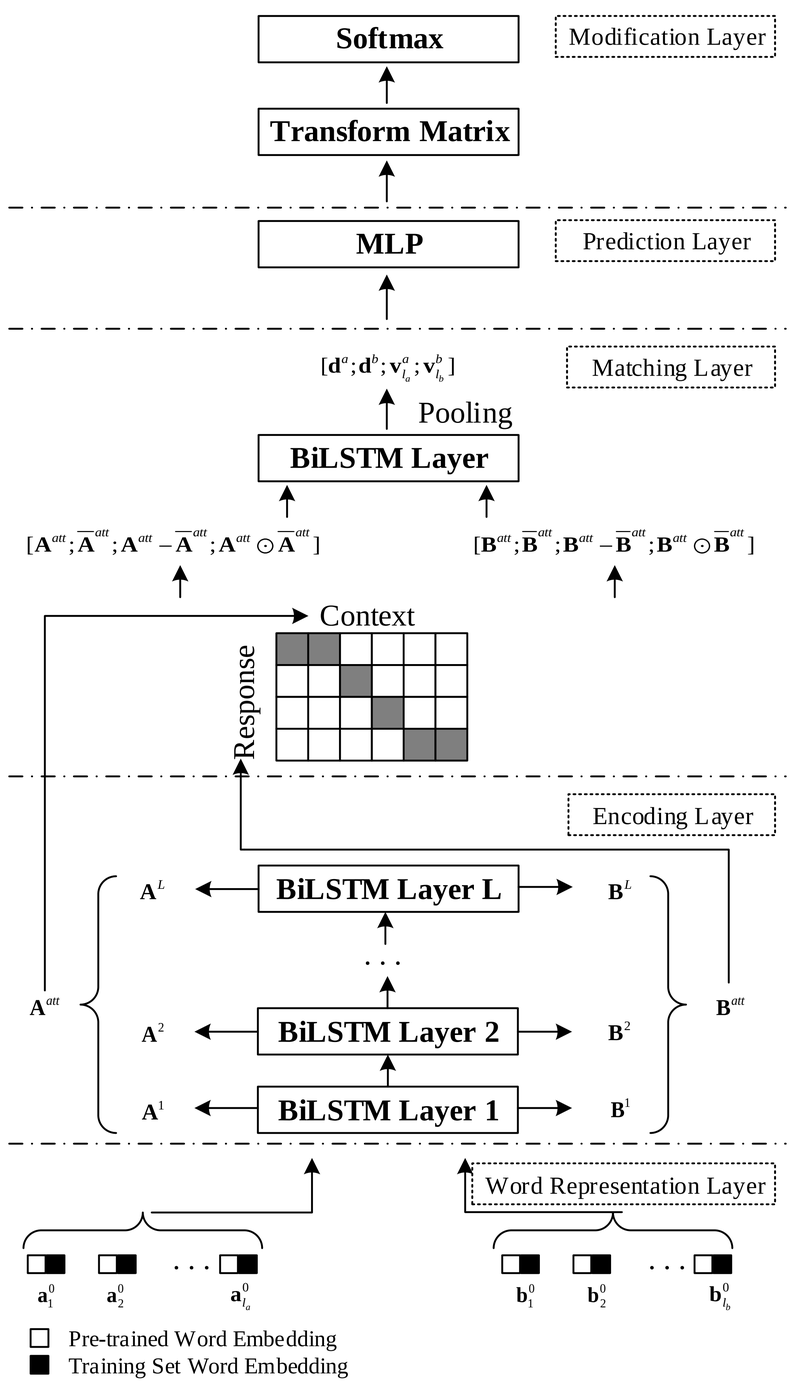
- One is the Advising dataset which is focused and small.
- The other one is the Ubuntu dataset which is more diverse and large.

The task is divided into 5 subtasks. A participant may participate in one, several, or all the subtasks to meet different goals for different subtasks. We only participate in the subtask 1 of this track, which aims to select the next utterance from **a candidate set of 100 utterances**.

Methodology

We propose a sequential model for end-to-end response selection by improving the original ESIM model from the following four aspects.

- A new word representation method which **combines the general pre-trained word embeddings with those estimated on the task-specific training set** to address the challenge of OOV words.
- An attentive hierarchical recurrent encoder (AHRE) is designed which **encodes sentences hierarchically and generates sentence representations by aggregating with attention**.
- A new pooling method which combines **multidimensional pooling and last-state pooling**.
- A modification layer is added before the softmax layer to **emphasize the importance of the last utterance in the context** for response selection.



Results

The results of our model on Ubuntu dataset and Advising dataset are summarized as follows. In the released evaluation results of DSTC7, our proposed method **ranked second on the Ubuntu dataset and third on the Advising dataset in subtask 1 of Track 1.**

Dev/Test	Dataset	R_{100}@1	R_{100}@10	R_{100}@50	MRR
Dev	Ubuntu(single)	0.521	0.817	0.982	0.616
	Ubuntu(ensemble)	0.534	0.825	0.982	0.631
	Advising(single)	0.206	0.556	0.906	0.323
	Advising(ensemble)	0.260	0.626	0.930	0.377
Test	Ubuntu(ensemble)	0.608	0.853	0.984	0.691
	Advising-1(ensemble)	0.420	0.766	0.972	0.538
	Advising-2(ensemble)	0.194	0.582	0.908	0.320

Analysis

We further investigated the effects of different parts in our proposed model by removing them one by one built on the Ubuntu development dataset.

	R_{100}@1	R_{100}@10	R_{100}@50	MRR
Our model (single)	0.521	0.817	0.982	0.616
-Modification layer	0.514	0.804	0.981	0.611
-Attentive hierarchical recurrent encoder	0.506	0.799	0.977	0.602
--Multi-dimentional and last-state pooling	0.500	0.791	0.974	0.598
--Fixed word embedding	0.488	0.776	0.969	0.591

Conclusion

In this paper, we have introduced our end-to-end model proposed for the response selection task in DSTC7. This model improves the original ESIM model from several aspects:

- Enhanced word representations
- AHRE for sentence encoding
- Multi-dimentional and last-state pooling for context-response matching
- Score calculation with emphasis on the last utterance in the context.

Our proposed method ranked second on the Ubuntu dataset and third on the Advising dataset in subtask 1 of Track 1. Our future work includes to design a more domain-general framework that can alleviate domain-dependency of models.

Reference

Chen Q, Zhu X, Ling Z H, et al. Enhanced LSTM for Natural Language Inference, ACL' 17

Dong J, Huang J. Enhance word representation for out-of-vocabulary on Ubuntu dialogue corpus, arXiv' 18

Shen T, Zhou T, Long G, et al. Disan: Directional self-attention network for rnn/cnn-free language understanding, AAAI' 18

Contact

Email: gujc@mail.ustc.edu.cn
Github: https://github.com/JasonForJoy/DSTC7-ResponseSelection