A Comparison between CNN and BiLSTM Models for Sequence Labeling

Zeqiang Lai
Department of Computer Science
Beijing Institute of Technology
1120161865@bit.edu.cn

Wenzhuo Liu
Department of Computer Science
Beijing Institute of Technology
1120161868@bit.edu.cn

Jinxuan Jin
Department of Computer Science
Beijing Institute of Technology
1120161864@bit.edu.cn

Tian Huan
Department of Computer Science
Beijing Institute of Technology
1120161861@bit.edu.cn

Anteng Li
Department of Computer Science
Beijing Institute of Technology
1120161866@bit.edu.cn

Xueyan Guo
Department of Computer Science
Beijing Institute of Technology
1120162336@bit.edu.cn

Abstract

In this paper, we replicate the result of De-CNN[2] for aspect extraction and perform several experiments with different configurations. For evaluation, we examine the results of De-CNN with different settings and the results of De-CNN and BiLSTM. We also make an attempt to apply a similar structure to named entity recognition task and carry out the same evaluation.

1 Introduction

Many natural language processing tasks can be converted into sequence labeling tasks, such as aspect term extraction and named entity recognition.

In Named Entity Recognition, given a sentence, we need to give a tag to each word. The tags look like ”B-LOC, I-LOC, B-PER, I-PER, O,...” where ”B” and ”I” indicate the beginning word, intermediate word of entities, ”LOC, PER,...” are the categories of entities and ”O” represents words that are not entities.

In Aspect Term Extraction, similar strategy is applied. We use ”B” to tag the beginning word of an aspect term. ”I” is used to tag the intermediate words. Other words that are not belong to any aspect term are tagged by ”O”.

2 Experiments

2.1 Datasets

For aspect extraction, we use semeval annual competition(2014)'s datasets[1]. The origin datasets are for Aspect Based Sentiment Analysis which including four subtasks(Aspect term extraction, Aspect

[1]http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools

term polarity, Aspect category detection, Aspect category polarity). We use python script to generate the datasets for aspect extraction and spilt them into three part, training, validation and testing sets.

For named entity recognition, we use CoNLL-2003 datasets[1] that contain independent English entity labels for English. The dataset contain four different types of named entities: locations, persons, organizations, and miscellaneous entities that do not belong in any of the three previous categories.

2.2 Aspect Extraction

Table2.2 presents the comparisons between De-CNN, it's variants and BiLSTM model for Aspect Extractions task. Scores of the models with both restaurants and laptop datasets are reported for completeness.

Since De-CNN has several components that can be tweak to understand their impact to the overall performance, we explore the impact that different convolutional layer, domain embedding and dropout has on De-CNN model. As we can see in the table, fewer number of convolutional layers gives a slightly decrease on the performance. Dropout has a significant improvement on the performance of +2.5 in F_1. This result is in line with our expectation, because the datasets is relatively small and the model tend to overfit to training data without dropout.

What surprises us is that there is dramatic increase in F_1 score[2](17.5 for restaurants and 22.8 for laptop) when we skip the domain embedding and only use general embedding. The results are incredible and we haven't come up with an reasonable explantation yet. We are planning to do more researches on it in the future.

The results of Bi-LSTM are also shown in the table.

Model	Variant	F1(Restaurant)	F1(Laptop)
De-CNN	5conv(full) + dropout + domain	79.7	74.9
De-CNN	5conv(full) + domain	77.2	71.0
De-CNN	4conv + dropout + domain	78.3	74.1
De-CNN	3conv + dropout + domain	79.1	73.3
De-CNN	2conv + dropout + domain	79.1	72.8
De-CNN	5conv(full) + dropout	97.2	97.7
De-CNN	5conv(full)	96.6	97.3
BiLSTM	origin	78.29	62.63

Table 1: Aspect Extraction results with De-CNN and BiLSTM, using different configurations. "5conv(full)","4conv", "3conv" and"2conv" refer to the models with different number of convolutional layers. "dropout" refer to all dropout layers. "domain" refer to the domain embedding.

2.3 Named Entity Recognition

We also experiment De-CNN in Named Entity Recognition task. Talbe2 shows the results. As we can see, the performance of BiLSTM is better than De-CNN with an increase of 10.7 in F_1 score.

We think the origin structure might not perfectly fit the NER task, because entity recognition need to distinguish the categories of entities and thus is more difficult than aspect extraction. Besides, the datasets might have an impact on the performance. The datasets for entity recognition is larger than the datasets for aspect extraction in our experiments.

[2]Our test set is different the De-CNN's origin paper[2], it might not be comparable with the results in it.

Table 2: NER results with De-CNN and BiLSTM

Model	Variant	F1
De-CNN	5conv(full)+dropout	62.4
BiLSTM	origin	73.1

3 Conclusion

In our experiment, we replicate the results of different types of De-CNN model and make an attempt to apply De-CNN on entity recognition task. By comparison, we prove that dropout is a good way to prevent overfitting in CNN models. We also have a surprising results on aspect extraction by removing the domain embedding.

The development in this paper results in some fruitful directions for future research. First, the impact of domain embedding on the performance is worthwhile to be further analyzed. Second, the origin De-CNN model could be reconstruct to better fit the entity recognition task. Finally, the question ”is it possible to combine Bi-LSTM and De-CNN to get better result?” need to be answered.

References

[1] Erik F Tjong Kim Sang and Fien De Meulder. “Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition”. In: *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics. 2003, pp. 142–147.

[2] Hu Xu et al. “Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction”. In: *arXiv preprint arXiv:1805.04601* (2018).