# Interactive Neural Network: Leveraging Part-of-Speech Window for Aspect Term Extraction (Student Abstract)

**Da Yin, Xiuyu Wu, Baobao Chang**
The MOE Key Laboratory of Computational Linguistics, Peking University
{wade_yin9712, xiuyu_wu,chbb}@pku.edu.cn

### Abstract

Aspect term extraction is a fundamental task for aspect-level sentiment analysis. Previous methods tend to extract noun aspect terms due to the large quantities of them, and perform badly on extracting aspect terms containing words with other POS tags, according to experimental results. In addition, few works focus on the POS tags of adjacent words which are critical to aspect term extraction. We propose a novel model which combines POS and word features in an interactive way, and makes full use of the POS tags of adjacent words by POS window. We conduct experiments on two datasets, and prove the effectiveness of our model.

## Introduction

Aspect term extraction (ATE) aims to identify the aspect terms present in the sentence. For example, given a sentence *'It has a good screen but the hard disk is very noisy.'*, models should return two aspect terms: *screen* and *hard disk*. Models are asked to label each word by one of $\{B, I, O\}$. Intuitively, Part-of-Speech (POS), which is not considered in most previous works, is essential to extracting aspect terms. The POS tags of aspect terms tend to be *noun* rather than *verb*, *article*, etc. Some works concatenate POS and word features, and show POS is useful. However, there is a main problem. Some aspect terms contain words with other POS tags, such as *verb* and *adjective*. For example, in the sentence *'It is easy to use.'*, while the word *'use'* is a *verb*, it is also an aspect term. Due to the large quantities of noun aspect terms, compared with nouns, the words without POS tag *noun* might have a lower probability of being detected as an aspect term. Previous baselines which leverage vanilla POS information have a tendency to extract noun terms. But for the words without POS tag *noun*, these methods tend to neglect them, and perform badly.

The POS tags of adjacent words are key to ATE. First, the POS tags of adjacent words are great supplement for individual POS tag information. The model would focus on the pattern composed by the POS information of a word and its surrounding words. For instance, for the word *"works"* in *"The computer works well."*, the POS information is transferred from *VBZ* to the patterns like [*NN, VBZ, RB*]. Compared with *VBZ*, [*NN, VBZ, RB*] is a more common pattern which symbolizes the existence of an aspect term. In addition, the POS tags of adjacent words provide a clue for the system to accurately label the sequence. For instance, if the word before a *noun* is an *adjective* and the next word is another *noun*, we may speculate the current word is the beginning of an aspect term. To take advantage of POS information for ATE task, we use the POS window (**PW**), which contains the POS information of each word and its adjacent words.

According to the good performance of leveraging vanilla POS, we can infer that the performance will be improved by incorporating PW into word features. Moreover, we propose a novel model, called Interactive Neural Networks (**INN**), which further builds the interaction between PW and word features. We conduct experiments on two datasets, and our proposed model achieves state-of-the-art performance on both of them. We also show the generality of PW from the fact that it can improve the performance of existing models.

## Model

The word embeddings of input sentence are $\mathbf{W} = [w_1, ..., w_i, ..., w_n]$ where $w_i \in \mathbb{R}^{d_w}$. The $i$-th word's POS Window $pw_i$ is $[pos_{i-k}, ..., pos_i, ...pos_{i+k}]$, where $2k+1$ is the size of PW. Then, the PW information of the $i$-th word is transformed into a low-dimensional vector $p_i$, where $p_i \in \mathbb{R}^{d_p}$. The **W** and **PW** make up the inputs of our model.

First, INN utilizes Bi-LSTM to learn the representations for word and PW, and the context-aware representations are $\text{hw}_i$ and $\text{hp}_i$ for word and PW, respectively. The hidden state is the sum of forward and backward LSTM. The reason we apply Bi-LSTM to PW is that there exists dependencies between PWs of adjacent words because the ranges of adjacent PWs overlap each other.

The crucial part of INN is an interactive layer. The layer integrates word and PW information:

$$\text{hw}_i^1 = \tanh(W_{pw}\text{hp}_i + \text{hw}_i), \quad (1)$$

$$\text{hp}_i^1 = \tanh(W_{wp}\text{hw}_i + \text{hp}_i). \quad (2)$$

| Method | Laptop | Restaurant |
|---|---|---|
| Bi-LSTM* | 75.25 | 71.26 |
| Bi-LSTM+POS* | 76.47 | 72.43 |
| MIN* | 77.58 | 73.44 |
| HAST | 79.52 | 73.61 |
| GloVe-CNN* | 77.67 | 72.08 |
| DE-CNN* | 81.59 | 74.37 |
| INN+POS* | 80.10 | 74.46 |
| INN+PW3-w/o-Interact* | 80.28 | 73.97 |
| INN+PW5* | 80.31 | 74.20 |
| INN+PW3-w/o-CRF* | 80.54 | 74.11 |
| Bi-LSTM+PW3* | 77.67 | 73.51 |
| GloVe-CNN+PW3* | 79.02 | 74.08 |
| DE-CNN+PW3* | 81.67 | 74.86 |
| INN+PW3* | $81.78^{\sharp}$ | $75.67^{\sharp}$ |
| INN+PW3+DE* | $\mathbf{82.24}^{\sharp}$ | $\mathbf{75.69}^{\sharp}$ |

Table 1: The performance (F1:%) of baselines and our model. The number behind PW denotes its size. The model with $*$ means its result is the average value of 5 runs. The result with $\sharp$ means statistical significant at the level of 0.05.

Based on the interacted representations of word and PW, the other two Bi-LSTMs are used to learn deeper representations for them and the results are $\text{hw}_i^2$ and $\text{hp}_i^2$. Both $\text{hw}_i^2$ and $\text{hp}_i^2$ of the $i$-th word are regarded as the final representation $r_i = [\text{hw}_i^2, \text{hp}_i^2]$. $[,]$ means concatenation. With a fully-connected layer, $r_i$ is projected to the label space $y_i$, of which the dimension is $d_a$. In the end, we will use CRF to model the dependencies between labels.

## Experiments

### Setup

To evaluate the performance of our proposed model, we do experiments on two datasets: laptop domain in SemEval 2014 Task 4 (Pontiki et al. 2014) and restaurant domain in SemEval 2016 Task 5 (Pontiki et al. 2016). The word embeddings are initialized by GloVe embeddings, and PW embeddings are randomly initialized. The size of PW is set to 3. For each run, We randomly select 150 samples from training data as the development data and utilize early stopping on the development set, and use the averaged results of 5 runs. To evaluate our model, we compare it with 6 baselines (Li and Lam 2017; Li et al. 2018; Xu et al. 2018). Besides, we conduct some experiments to verify the effectiveness of each part in our model: INN+POS uses the vanilla POS. INN+PW3-w/o-Interact directly concatenates the representations of PW with size of 3 and word in the end without interaction. INN+PW5 uses PW with size of 5. INN+PW3-w/o-CRF does not use CRF. To verify the generality of PW, we construct three baselines: Bi-LSTM+PW3, GloVe-CNN+PW3 and DE-CNN+PW3.

### Results and Analysis

Table 1 shows the performance of all the methods described above. We observe that our model INN+PW3 achieves the state-of-the-art performance on both datasets compared with the best baseline DE-CNN. We also apply domain embeddings in DE-CNN to ours, and the model INN+PW3+DE further improves the performance. Then, we evaluate the effectiveness of POS and INN structure. We observe that Bi-LSTM+POS outperforms Bi-LSTM, suggesting POS is critical to ATE. Additionally, INN+POS greatly improves over Bi-LSTM+POS. This proves that INN performs much better in leveraging POS. Moreover, we evaluate the effectiveness of PW3. INN+PW3 outperforms INN+POS on two datasets, and Bi-LSTM+PW3 also improves the performance compared with Bi-LSTM+POS. As shown in Table 2, we observe that the models with vanilla POS perform badly in extracting words with non-*noun* POS tags. Replacing POS with PW3, a substantial improvement could be observed. Also, we notice that INN+PW3 performs better than INN+PW5. The reason is that the quantity of PW5 is much larger than that of PW3, and this probably leads to data sparsity problem. Thus, it is necessary to select a proper size for PW. Finally, we find that with PW3, the performance is improved greatly compared with the corresponding base models. This means that PW is broadly applicable and beneficial.

| Method | Laptop | Restaurant |
|---|---|---|
| Bi-LSTM+POS | 62.34 | 49.75 |
| Bi-LSTM+PW3 | 69.16 | 54.89 |
| INN+POS | 67.49 | 51.73 |
| INN+PW3 | 72.10 | 58.70 |

Table 2: The performance (F1:%) of extracting the aspect terms containing words with not *noun* POS tags. The proportions of these terms in two sets are about 26% and 18%.

## Conclusion and Acknowledgement

We propose a new method to improve the performance of ATE by making full use of POS information. Experimental results show that our model achieves state-of-the-art performance on two datasets.

We would like to thank the anonymous reviewers for the helpful discussions and suggestions. This work is supported by National Natural Science Foundation of China under Grant No.61876004, No.61751201 and M1752013. The corresponding author of this paper is Baobao Chang.

## References

Li, X., and Lam, W. 2017. Deep multi-task learning for aspect term extraction with memory interaction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2886–2892.

Li, X.; Bing, L.; Li, P.; Lam, W.; and Yang, Z. 2018. Aspect term extraction with history attention and selective transformation. *arXiv preprint arXiv:1805.00760*.

Pontiki, M.; Galanis, Dimitris Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; and Manandhar, S. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval-2014)*, 19–30.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; Mohammad, A.-S.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 19–30.

Xu, H.; Liu, B.; Shu, L.; and Yu, P. S. 2018. Double embeddings and cnn-based sequence labeling for aspect extraction. *arXiv preprint arXiv:1805.04601*.