Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction

Hu Xu¹, Bing Liu¹, Lei Shu¹, Philip S. Yu¹,²

¹Department of Computer Science, University of Illinois at Chicago
²Institute for Data Science, Tsinghua University

{hxu48, liub, lshu3, psyu}@uic.edu

Aspect Extraction is an important task in fine-grained sentiment analysis. We propose a simple and fast approach without using any sophisticated features and models.

The contributions are in 2 folds:
- **Double embedding**: we use two types of pre-trained embeddings for aspect extraction: general purpose embeddings and domain specific embeddings.
- **CNN**: we use CNN for sequence labeling, which is parallel and faster than serial LSTM. We adapt CNN (e.g., drop max-pooling layer) to get better results.

In a Laptop review:

- **B**
- O
- O
- O

Its speed is incredible

Laptop Embedding

General Embedding (GloVe)

Speed means how many instructions (not meters) per second in laptop

The r-th CNN filter for the i-th word in layer l

\[ x_{i,r}^{(l+1)} = \max \left( 0, \left( \sum_{j=-c}^{c} w_{j,r}^{(l)} x_{i+j}^{(l)} + b_{r}^{(l)} \right) \right), \quad (1) \]

Table 1: Dataset description with the number of sentences(#S.) and number of aspect terms(#A.)

<table>
<thead>
<tr>
<th>Description</th>
<th>Training #S./#A.</th>
<th>Testing #S./#A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SemEval-14 Laptop</td>
<td>3045/2538</td>
<td>800/654</td>
</tr>
<tr>
<td>SemEval-16 Restaurant</td>
<td>2000/1743</td>
<td>676/622</td>
</tr>
</tbody>
</table>

Table 2: Comparison results in F₁ score: numbers in the third group are averaged scores of 5 runs. * indicates the result is statistical significant at the level of 0.05.

The implementation of the paper is available at https://www.cs.uic.edu/~hxu/