**Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction**

1 Introduction

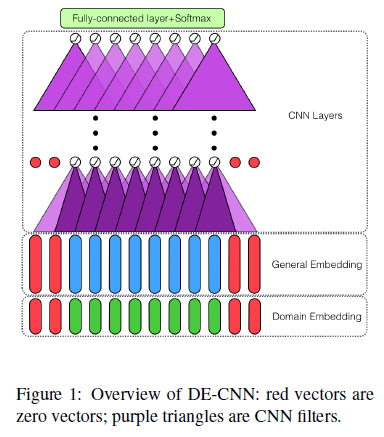
Aspect extraction is an important task in sentiment analysis (Hu and Liu, 2004) and has many applications (Liu, 2012). It aims to extract opinion targets (or aspects) from opinion text.

To address the first consideration, we propose a double embeddings mechanism that is shown crucial for aspect extraction. Thus, we leverage both general embeddings and domain embeddings and let the rest of the network to decide which embeddings have more useful information.

To address the second consideration, we use a pure Convolutional Neural Network (CNN) model for sequence labeling.

2 Related Work

3 Model



4 Experiments

4.1 Datasets

4.2 Baseline Methods

4.3 Hyper-parameters

4.4 Results and Analysis

The double embedding mechanism improves the performance and in-domain embeddings are important. DE-CNN has two major types of errors. One type comes from inconsistent labeling (e.g., for the restaurant data, the same aspect is sometimes

labeled and sometimes not). Another major type of errors comes from unseen aspects in test data that require the semantics of the conjunction word “and” to extract. For example, if A is an aspect and when “A and B” appears, B should also be extracted but not. We leave this to future work.

5 Conclusion

We propose a CNN-based aspect extraction model with a double embeddings mechanism without extra supervision. Experimental results demonstrated that the proposed method outperforms state-of-the-art methods with a large margin.