Neural Aspect and Opinion Term Extraction with Mined Rules as Weak Supervision (Appendix)

Hongliang Dai
Department of CSE, HKUST
hdai@cse.ust.hk

Yangqiu Song
Department of CSE, HKUST
yqsong@cse.ust.hk

A Experimental Results with BERT

We also conduct experiments with the language representation model BERT. The original paper suggests two approaches to apply this model to a sequence labeling problem. The first approach is to fine tune BERT by feeding the final hidden representation for each token into a classification layer; the second approach is to use the hidden layers of the pretrained Transformer as fixed features for a sequence labeling model. They are called the fine-tuning approach and the feature-based approach, respectively. Obviously, the feature-based approach can also be used by RINANTE.

Here, we compare four different appraoches: BiLSTM-CRF + word2vec, BERT fine-tuning, BERT feature-based and RINANTE+BERT. BiLSTM-CRF + word2vec simply uses word2vec embeddins as the input of a BiLSTM-CRF. BERT fine-tuning is the same fine-tuning approach for name entity recognition used in the paper that proposes BERT. BERT feature-based uses the extracted features as the input of a BiLSTM-CRF model. RINANTE+BERT uses our approach RINANTE-DOUBLE-Pre, but with the features extracted by BERT as word embeddings. Both BERT feature-based and RINANTE+BERT use the top four hidden layers as features.

We use the pretrained *BERT-Base, Uncased* model and further pretrain it with the Yelp reviews and Amazon reviews for our restaurant datasets and laptop dataset, respectively. 200-dimensional BiLSTMs are used for both BERT feature-based and RINANTE+BERT.

The experimental results are listed in Table 1. We can see that using BERT yields better performance than using word2vec. RINANTE is still able to further improve the performance when contextual embeddings obtained with BERT are used.

	SE14-R		SE14-L		SE15-R	
Approach	Aspect	Opinion	Aspect	Opinion	Aspect	Opinion
BiLSTM-CRF + word2vec	84.06	84.59	73.47	75.41	66.17	68.16
BERT fine-tuning	84.36	85.50	75.67	79.75	65.84	74.21
BERT feature-based	85.14	85.74	76.81	81.41	66.84	73.92
RINANTE+BERT	**85.51**	**86.82**	**79.93**	**82.09**	**68.50**	**74.54**

Table 1: Aspect and opinion term extraction performance.