Supplementary Material of "Commonsense LocatedNear Relation Extraction"

1 Features for SVM classifier

The 6 types of features used in SVM are as follows:

1. *Bag of Words (BW)* The set of words that ever appeared in the sentence.
2. *Bag of Path Words (BPW)* The set of words that appeared on the shortest dependency path between objects e_i and e_j in the dependency tree of the sentence s, plus the words in the two subtrees rooted at e_i and e_j in the parse tree.
3. *Bag of Adverbs and Prepositions (BAP)* The existence of adverbs and prepositions in the sentence as binary features.
4. *Global Features (GF)* The length of the sentence, the number of nouns, verbs, adverbs, adjectives, determiners, prepositions and punctuations in the whole sentence.
5. *Shortest Dependency Path Features (SDP)* From the dependency parse tree of the sentence and the shortest path between the two objects e_i and e_j.
6. *Semantic Similarity Features (SS)* The cosine similarity between the pre-trained GloVe word embeddings (Pennington et al., 2014) of the two object words.

2 Choices of f scoring function

In the stage of LocatedNear relation extraction from corpus, we have to choose a scoring function to aggregate the results from sentence-level relation classifier. After calculating the probability score from LSTM+Norm classifier, we could compare the performance of our 5 choices of scoring function f:

$$f_0 = m \tag{1}$$

$$f_1 = \sum_{k=1}^{m} \text{conf}(s_k, e_i, e_j) \tag{2}$$

$$f_2 = \frac{1}{m} \sum_{k=1}^{m} \text{conf}(s_k, e_i, e_j) \tag{3}$$

$$f_3 = \sum_{k=1}^{m} 1_{\{\text{conf}(s_k,e_i,e_j)>0.5\}} \tag{4}$$

$$f_4 = \frac{1}{m} \sum_{k=1}^{m} 1_{\{\text{conf}(s_k,e_i,e_j)>0.5\}} \tag{5}$$

We use each of the scoring functions to rank the 500 commonsense LocatedNear object pairs described in Section 3 of the paper. Table 1 shows the ranking results using Mean Average Precision (MAP) and Precision at K as metric. Accumulative scores (f_1 and f_3) generally do better. Thus we choose $f = f_3$ as our scoring function throughout the paper.

f	MAP	P@50	P@100	P@200	P@300
f_0	0.42	0.40	0.44	0.42	0.38
f_1	0.58	**0.70**	0.60	0.53	**0.44**
f_2	0.48	0.56	0.52	0.49	0.42
f_3	**0.59**	0.68	**0.63**	**0.55**	**0.44**
f_4	0.56	0.40	0.48	0.50	0.42

Table 1: Ranking performances of the 5 scoring methods.

References

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.