

Proposal 2 : Preposition Sense Disambiguation

Dhiraj Gurkhe
1209305002
dgurkhe@asu.edu

Himanshu Tyagi
1209283279
htyagi3@asu.edu

Sandeep Gunda
1208165552
sandeep.gunda@asu.edu

Shashank Singh
1208960398
ssing129@asu.edu

1 Problem Definition

In this project we propose a methodology to disambiguate prepositions in a given sentence/phrase. We want to find out the context in which the preposition was used to find its sense for the given data point. We plan to use supervised machine learning techniques with incorporation of NLP techniques to build a classification model. The classifier will classify the preposition in a given test example into its corresponding possible sense. Many models yet build like (Ye and Baldwin, 2007) use NLP techniques of POS tagging, chunking, parsing alike to make features for the classifier. We plan to do the same by using a combination of NLP parsing and common text processing techniques to leverage features of the given dataset. A baseline approach will be to build a bag of word model, and classify based on count of words. A Complex model will try to find the semantic relation between the set of words around the preposition, and make a mapping function for the various features. For the classification task various algorithm like navie bayes, SVM, decision trees could be experimented with to find the best fit. Deep learning with neural nets will also be test for the classification task. The challenge of the project lies in the two step process of appropriately parsing and mapping the preposition, and then to build a general feature extraction model for all the training data. As per our understanding we should be building n classification models, where n would be the number of distinct prepositions. The training and test data would be a spilt of the given dataset in the ratio of 2:1.

2 Motivation

Prepositions are a closed class, meaning that the number of prepositions remains relatively constant and that their meanings are relatively stable. De-

spite this, their treatment in computational linguistics has been somewhat limited. Prepositions are viewed as function words that occur with high frequency and therefore carry little meaning. Prepositions play an important role in task of semantic role labeling (Litkowski and Hargraves, 2007). When prepositions are considered in relation to words, they attain more importance. Most of the linguistic theory focuses on subjects and objects as the most important verb argument, but frequently preposition are also observed in such phrases and can't be ignored. With all being said, it can be seen that prepositions carry definite behaviour similar to the one carried by open classes. Therefore, when preposition are present in abundance, they can be treated to similar type of analyses as with open classes.

3 Methodology

3.1 Techniques

- **Classifier Models** : Navie Bayes, SVM, Random Forest, Neural Net.
- **Feature Engineering**: Part of speech tagging, Named entity recognition, Parsing, Conceptual Class extraction, Chunking, Word2vec modeling

3.2 Software Tools

- **NLP/feature Engineering** : Tensor Flow, NLTK, Gensim
- **Machine Learning** : Sklearn, keras

3.3 Datasets

- The preposition project dataset
- Framenet Database

4 Evaluation

The SemEval workshop presented participants with a formal preposition sense disambiguation task to encourage the development of systems for the disambiguation of preposition senses (Litkowski and Hargraves, 2007). The training and test data sets used for SemEval have been released to the general public, and we used these data test our system. The baseline is also provided using the most frequent and first sense method in the same paper.

References

- Ken Litkowski and Orin Hargraves. 2007. Semeval-2007 task 06: Word-sense disambiguation of prepositions. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 24–29. Association for Computational Linguistics.
- Patrick Ye and Timothy Baldwin. 2007. Melb-yb: Preposition sense disambiguation using rich semantic features. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 241–244. Association for Computational Linguistics.