

Homework 4: Project-Related Paper Report

Ankit Goyal
ankit@cs.utexas.edu
Natural Language Processing

April 22, 2015

Paper: Nazneen Fatema Rajani, Edaena Salinas and Raymond Money. *Using Abstract Context to Detect Figurative Language*

Review: We use metaphors in our everyday language either to help understand abstract or unfamiliar concepts, to express better with emotions or to draw attention. Understanding metaphors is an important enhancement to any NLP system since by understanding metaphors a system can better understand the natural language. The goal of the paper is to detect figurative language. In the paper, authors use rich combination of features to identify figurative language and is able to achieve state of the art results on the SemVal 2013 phrase semantics dataset.

Problem Statement: Understanding metaphors involves two steps: 1. detecting metaphors 2. understanding the meaning of metaphors. The above paper and our project considers only the problem of detecting metaphors.

Approach: The paper uses contextual features (discussed later) in a logistic-regression classifier to predict whether a phrase is being used literally or figuratively. The features used in the paper are:

1. **Bag of words:** Frequency of each word. This helps determine the usage of each word in a given context.
2. **Concreteness Measure:** A score assigned on the basis of degree of abstractness. For example, words like milk, heavy refer to concrete things whereas words such as politics, calculating refer to ideas and concepts that are more abstract. They use MRC Psycholinguist Database to get this score.
3. **Phrase to Context Relatedness:** This features represent the relatedness of the words to the given context. For example, the phrase *bread and butter*, occurring in literal sense would be related to *food* or *eating*. Whereas in figurative sense, it'll occur with less-related words. It's calculated using the WordNet similarity. All words package is used
4. **Topic to Context Relatedness:** For each phrase in the training phrase, topics were extracted for both literal and figurative sense. Relatedness of topic words was then calculated using the same method as phrase to context relatedness.

The data used was from SemEval 2013 task which contained phrase with both their literal use and figurative. They used LIBLINEAR's L2 regularized Logistic Regression (L2LR) Fan et al. (2008) to build classifiers using the extracted features.

They were able to obtain highest accuracy of 80% for the seen phrases and 64% for the unseen phrases.

Limitations: They were not able to achieve very good performance on the unseen phrases since as they mention that their features measured the abstractness only. They were able to do pretty well on the seen phrases though.

Our Approach: Above approach works well for known phrases whereas it doesn't have enough semantic information in the features to work on unseen phrases. We plan to extend the above approach by adding more semantic features and hope to achieve better results.

1. **Verb figurative likeliness:** We believe that the use of semantic role labelling could be a good candidate for this problem. Consider for example the phrase "My car drinks lots of Gasoline". In this example, the drinking by car is used as a metaphor for car using lots of Gasoline. Using the VerbNet, we can see that the verb drink requires it's Agent to be an animate object whereas "car" in the example phrase is not animate. To determine the roles of different words, for example, that the car is an inanimate object we plan to use FrameNet based semantic role labeler. First we use semantic role labelling using FrameNet to categorize the words in the phrase and then we check if selectional roles given by VerbNet are satisfied. If not then it's a good candidate for being a metaphor assuming that the phrase is grammatically correct.
2. **Selective weightage to special words:** Metaphors are more likely to contain words such as "like" or "as". So a higher weightage to the phrase could be given if any of these words is present.
3. **Metaphors are mostly in third person:** This is another feature that we'd like to try. We believe that most of the times metaphors are used when the subject is third person. Although the effectiveness of this feature will highly dependent on the corpus. But we believe it could be a good indicator whatsoever.
4. **Ontonotes features:** We also plan to use features from Ontonotes which relates to word sense ambiguation. Using this we could get a score based on how common the use of this word is in a given sense. We are still researching more on this feature.

Dataset: We plan to use the same dataset: SemEval 2013.

Conclusion: We believe that using semantic role labelling information, we'd be able to put restrictions on the verbs which metaphors usually tend to violate. Moreover we believe a feature as simple as giving more weightage to words like "as" and "like" will contribute significantly. Since the SemEval task, lot of work has been done in the metaphor identification which includes methods like verb-noun clustering, using bias of the verb, and other cognitive features. We hope to achieve good results using our features since they add more semantics to the feature set.

1 References:

1. Nazneen Fatema Rajani, Edaena Salinas and Raymond Money. *Using Abstract Context to Detect Figurative Language*
2. *The Second Workshop on Metaphor in NLP*. The Association for Computational Linguistics