Michael Meding
Hoanh Nguyen
13 April 2015
Natural Language Processing
Professor Anna Rumshisky

Progress Report

Project Code: https://github.com/mikemeding/SemEval-2015.git

Changes to the problem definition

As we are following Task 1 from SemEval 2015 our problem definition is fixed and has not changed

Additional literature you've looked at

All relevant literature used for this project can be found in the papers directory of the main project page.

- **Named Entity Recognition in Tweets An Experimental Study**, Alan Ritter, Sam Clark, Mausam and Oren Etzioni
- **Extracting Lexically Divergent Paraphrases from Twitter**, Wei Xu, Alan Ritter, Chris Callison-Burch, William B Dolan, and Yengfeng Ji
- **Paraphrase Identification on the Basis of Supervised Machine Learning Techniques**, Zornitsa Kozareva and Andres Montoyo
- **Data-Driven Approaches for Paraphrasing Across Language Variations**, Wei Xu
- **Extracting Lexically Divergent Paraphrases from Twitter**, Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan and Yangfeng Ji

The experiments and evaluation you have conducted so far

- Reimplemented Python logistic regression in Java. Slightly modified base feature extractor. Made use of some lexical resources which we had from previous experiments.
- Sentiment does not appear to be good for paraphrasing Tweets. This will be tested more in depth but the initial results are underwhelming.
- Our plan now is to implement as many features as possible to improve our model and maximize our results before the end of this project.