CS447 Project Idea - Detecting Stance in Tweets

Bhargav Mangipudi (mangipu2), Vishaal Mohan (vmohan9)

Fall 2015

This project will focus on the problem of detecting stance in tweets. This is one of the tasks defined for the SemEval 2016 conference. The task defines the problem statement as we define stance detection to mean automatically determining from text whether the author is in favor of the given target, against the given target, or whether neither inference is likely.

One example is for the target is

- Target: legalization of abortion

- Tweet: A foetus has rights too! Make your voice heard.

The stance analysis would classify the tweet as against the given target.

There has been some research into classifying stance in debates [1, 2] and online texts[3]. This project involves an application of sentiment analysis to the tweet medium. The 140-character limit and the implied brevity of speech increases the challenge here. Another important feature is the use of hashtags in tweets for representing a user's stance.

Our idea is to look at different sets of representation of the tweet content using the simple bag-of-words approach, the word vector approach [4, 5, 6] and other related approaches. We would like to do a comprehensive study of different approaches to the problem and their relative performance on the test data.

Since the contest submission is in the month of Jan 2016, if we get to a good state by the end of the course, we plan to tune the work and submit it to the contest.

Link to the task page: `http://alt.qcri.org/semeval2016/task6/`
Link to available dataset: `http://alt.qcri.org/semeval2016/task6/index.php?id=data-and-tools`

SemEval 2016 - Task Timeline for reference
Trial data ready - June 30, 2015
Training data ready - August 30, 2015

Test data ready - December 15, 2015
Evaluation start - January 10, 2016
Evaluation end - January 31, 2016
Paper submission due - February 28, 2016 [TBC]
Paper reviews due - March 31, 2016 [TBC]
Camera ready due - April 30, 2016 [TBC]
SemEval workshop - Summer 2016

References

[1] Kazi Saidul Hasan and Vincent Ng. Stance Classification of Ideological Debates: Data, Models, Features, and Constraints. In *IJCNLP*, pages 1348–1356, 2013.

[2] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M. Mohammad. Sentiment Analysis of Short Informal Texts. *J. Artif. Intell. Res. (JAIR)*, 50:723–762, 2014.

[3] Rob Abbott Jean E. Fox Tree Robeson Bowmani Michael Minor Pranav An, Marilyn Walker. Cats Rule and Dogs Drool!: Classifying Stance in Online Debate.

[4] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. *Proceedings of NAACL-HLT*, (June):746–751, 2013.

[5] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. *Nips*, pages 1–9, 2013.

[6] Tomas Mikolov, Greg Corrado, Kai Chen, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, pages 1–12, 2013.