# NLP Project Paper

Ashwini Venkatesh
University of Texas at Austin
ashuven6@cs.utexas.edu

## 1. Problem Description

In this work, the authors propose an approach for stance detection in context of online debates [1]. Stance detection is different from sentiment analysis since in this task we now have a target topic or issue. A piece of text suggests whether the author is in favor, against or neutral towards the target.

## 2. Motivation

Understanding the stance in conversations can be useful. We can leverage this information to generate summaries for debates, understanding what makes an argument successful and persuasive, identifying the linguistic reflexes of perlocutionary acts such as persuasion and disagreement. In this paper the aim is to automatically identify rebuttals and further identify the author's stance towards a particular target.

## 3. Corpus

The corpus used consists of 1113 two-sided debates (4873 posts) from Convinceme.net for 12 topics ranging from casual debates such as Firefox vs. IE to more issue related topics such as Capital Punishment, Abortion. In total the corpus consists of 2,722,340 words; the topic labeled debates used in the experiments contain 507,827 words. The authors have analyzed the data for each topic in terms of no of posts, no of rebuttals, posts per author, percentage of authors with more than one post, post length, positive and negative emotion words and Words per sentence. Human annotators are used to identify rebuttals and non-rebuttals and determine the top human accuracy on these tasks. This is done using Amazon Mechanical Turk and is shown that Rebuttals were clearly harder to side: annotators correctly sided non-rebuttals 87% of the time, but only managed 73% accuracy for rebuttals.

## 4. Methodology

Evaluation is done by training two classifiers with different properties: NaiveBayes and JRip. JRip is a rule based classifier which produces a compact model suitable for human consumption and quick application. Following features are used to train the model :

- Counts like post length, Unigrams and Bigrams
- Cue Words using initial unigram, bigram and trigram sequences to capture the usage of cue words to mark responses of particular type, such as oh really, so, and well. These are typically present in online blog posts.
- Repeated Punctuation like !!, ?? which sometimes captures the emotion of the author.
- Syntactic Dependency where a dependency structure to determine the target of an opinion word.
- Generalized dependencies, where you back off the head word in each of the above features to its part-of-speech tag
- Context features where you examine the features of the parent post. This can be used since the structure is such that the posts are a reaction to a previous post.
- LIWC. We also derived features using the Linguistics Inquiry Word Count tool. LIWC provides meta level conceptual categories for words to use in word counts. Some LIWC features that are computed are words per sentence (WPS), pronominal forms (Pro), and positive and negative emotion words (PosE) and (NegE).

## 5. Key Takeaways

The paper evaluates Rebuttal Classification and Automatic Debate-Side Classification. In case of rebuttal classification the learned rules using JRip were almost entirely based on LIWC and unigram lexical features, such as 2nd person pronouns, quotation marks, question marks, and negation, all of which correlated with rebuttals. They evaluate Automatic debate-side classification using naive bayes both with context and without context. The LIWC feature set is the only feature set that appears to show improvement over a set of topics. This is probably because this feature set is based on a lexical hierarchy that includes social features, negative and positive emotion, and psychological processes. Human agreement on this subject is less as well. They show that ideological debates feature a greater share of rebuttal posts, and that rebuttal posts are significantly harder to classify for stance, for both humans and trained classifiers.

## 6. Strength and Limitations of the paper

The authors seem to have experimented with a wide range of features. However the empirical results presented show that they are not very successful in leveraging these features. They conclude that Naive Bayes classifier is better than rule based classifier while evaluating automatic debate-side classification but provide no insight as to why this happens. Further the the paper doesn't present a very clear analysis of different feature sets. A detailed analysis of why some features work and matter in predicting the final value is minimal. They do not provide an reasons to support the experimental data obtained.

## 7. Project

In our project we aim to solve the task of stance detection in tweets. In this task we look at tweets to predict their stance as opposed to the paper discussed above where the corpus contains posts from online debates. This task is part of the International Workshop on Semantic Evaluation SemEval-2016. They provide training dataset in the form of tweets labelled with the target and stance. The challenge lies in the fact that tweets have less textual information compared to the data set used by [1]. Thus we have very less information or structure to work with and predict the stance. A tweet is limited to 140 characters. Also the language used in tweets is very informal and there are additional elements like hash tags in a tweet as opposed to traditional blog posts.

Similar to the paper we will use NGrams as one the features for training. As stated by [4] the unigram baseline can be difficult to beat for certain types of debates. We plan to train the model using unigrams and bigrams. Apart from this we we also use dependency parse trees to determine the subject, opinion, and whether a negator affects a word in the graph.

The corpus discussed above has posts which are long. The average length of the posts are 300 characters which is twice the number of characters in tweets. Unlike the paper discussed, we need to augment the tweet with some information about the target. Since a tweet is limited to 140 characters as opposed to a discussion forum, we need to build in some intuition about the target and what indicates a positive stance and negative stance in the issue or the target considered. Additional information could be introduced in the form of semantic relatedness of two topics using a Wikipedia link-based measured as proposed in [2]

Apart from this a tweet is not structured as a rebuttal as given in the paper. A rebuttal is a response to a particular post. Hence you know what the post is written in response to thus deriving some context from the parent. Raw tweets in our data set don't have any conversation structure.

[1] uses LIWC to capture the positive and negative emotion words. Instead we will replace this with a subjectivity and sentiment lexicons [3]. This will be much easier to compare the effect of using these lexicons as features as opposed to just the effect of LIWC. In my opinion, this would be one of the major features which can be used to learn the mapping.

One of the main features to consider in the task we are trying to solve is to leverage the hash tags in the step. The hash tags must first be pre-processed since these are usually concatenated phrases. However hashtags in many cases gives some valuable info because they essentially condense the essence of the tweet. For example in the tweet : Hilary Clinton surely doesn't seem to be a true New Yorker #bernieForPresident ! If the target is Bernie Sanders, we can clearly predict from the hash tag the author's polarity towards the target.

We plan to train an ensemble using the above mentioned features. In my opinion these features leverage all the aspects of a tweet which indicate stance. To compare the results obtained we plan to use a bag of words approach as the baseline. The bag-of-words approach might give a considerably good performance. However it is to be seen that by using the mentioned features what sort of improvement can we achieve on the baseline.

## References


[1] P. Anand, M. Walker, R. Abbott, J. E. F. Tree, R. Bowmani, and M. Minor. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '11, pages 1–9, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[2] A. Faulkner. Automated classification of stance in student essays: An approach using stance target information and the wikipedia link-based measure. In *The Twenty-Seventh International Flairs Conference*, 2014.

[3] E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 105–112. Association for Computational Linguistics, 2003.

[4] S. Somasundaran and J. Wiebe. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, pages 116–124, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.