Twitter sentiment blablabla

Yun Wu **Qiren Chen** **Xiaofan Lu**
University of Texas at Austin
Computer Science Department
{yun, qiren, xiaofan}@cs.utexas.edu

Abstract

This document contains the instructions for preparing a camera-ready manuscript for the proceedings of ACL-2014. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used for both papers submitted for review and for final versions of accepted papers. Authors are asked to conform to all the directions reported in this document.

1 Introduction

Microblogging websites (such as Twitter, Weibo) have gained popularity in recent years. People can easily post real time, short message to express their opinions. Sentiment of microblog is of particular interest because such information is valuable to both consumer and manufactures.

Most of existing works are based on bag of words classifiers. People propose and evaluate different features to improve the performance of bag-of-words model. Such classifiers can work well in longer documents by relying on a few words with strong sentiment such as ”great” or ”awesome”. However, such bag-of-words models have difficulties in handling in negation and comparisons. Negation and comparison involves the structure of sentence which clearly is not in the scope of bag-of-words models. To further improve the performance, new models which capture the structure of sentences are needed.

2 Related Work

3 Bag of Word Classifier

4 Recursive Neural Tensor Network

5 Pre-processing

5.1 Twitter related features

5.2 Slang Translation

5.3 Emoticons

5.4 Spelling Correction

5.5 Word Cluster

6 Experiment

6.1 Corpus

We conducted our experiment on three types of corpus.

- The first corpus is the Stanford sentiment treebank released by Socher et. al. (2013). It is based on the dataset introduced by Pang and Lee (2005) and consists of 11,855 single sentences extracted from movie reviews. It was parsed with the Stanford parser (Klein and Mannning, 2003) and includes a total of 215,154 unique phrases from those parse trees, each annotated by 3 human judges. It is not twitter message, but would give us a comparison of RNTN and SVM on well formatted English. We refer to this corpus by *Sentiment Treebank* in the reset of paper.

- The second corpus is movie reviews on Twitter, which can be divided into two categories.
  - The first categoriy is single tweet movie review taken from two specialized review accounts (@FilmReviewIn140, @MovieTwoosh). We have 364 tweets in this corpus. Such reviews are mostly well formatted, usually consist of several sentences. The author rated each movie with A to F grades. And if a movie receives a grade no worth than B, we label the review as positive, otherwise, it is negative. We refer to this corpus by *moiveA* in the reset of paper.
  - The second category of movie reviews are collected by searching two currently popular movies names (Rio2 & Captain American2). Such tweets are published by the generally public and they have all the noisy feature of tweet. We manually labeled this corpus. We refer to this corpus by *moiveB* in the reset of paper.

- The third corpus is general tweet message. It is taken from SemEval-2013: Sentiment Analysis in Twitter Task B[1]. Each of the tweet messages has been manually labeled as positive, negative, or neutral. Out of all the 5,750 messages, 2,042 are positive, 855 are negative and 2853 are neutral. We refer to this corpus by *SemEval* in the reset of paper.

[1] `http://www.cs.york.ac.uk/semeval-2013/task2/index.php?id=data`

6.2 Single Sentence Sentiment

We firstly evaluate both models using *Sentiment Treebank*, which contains single sentence movie reviews extracted from `http://www.rottentomatoes.com/`. We used the same training/testing splits as in the original paper by Socher et. al. (2013).

Model	Accuracy		
	positive	negative	overall
RNTN	80.83	87.91	84.27
SVM1			
SVM2			

Table 1: Binary decision

We also evaluated the performance of using the emotional label as a feature to train the SVM classifier. ..

6.3 Multiple Sentences Sentiment

We then evaluates how different models works decide the sentiment of the whole twitter message. This won't affect bag-of-word method much because now we only need a larger bag. However, RNTN relies on the structural of single sentence so we need to combine the sentiment from multiple sentences within a single tweet. Here, we use the model trained on *Sentiment Treebank* and tested on *movieA* corpus.

Model	Accuracy(%)		
	positive	negative	overall
RNTN_{hard}	70.08	81.54	74.18
RNTN_{soft}	78.21	80.0	78.85
SVM1			
SVM2			

Table 2: Binary decision

6.4 Effect of Preprocessing

We also evaluates how much preprocessing contributed to our final performance. In this experiment, we still use the model trained on *Sentiment Treebank* but tested on *movieB* corpus, which contains noisy common twitter message collected by searching movie names. We evaluated both models with and without preprocessing the original corpus.

Model	Overall Accuracy(%)	
	with pre-processing	w/o pre-processing
RNTN		
SVM1		
SVM2		

Table 3: Effect of preprocessing

Comparing both table, we can see that preprocessing indeed helped in improving the performance.

6.5 General topic Tweet

We conducted three sets of experiment over the general topic twitter corpus *SemEval*.

- Exp 1: Training on 90% of *SemEval* and testing on the rest 10%.

Model	Accuracy(%)		
	positive	negative	overall
RNTN			
SVM1			
SVM2			

Table 4: Experiment 1

- Exp 2: Training on *Sentiment Treebank* and testing on *SemEval*.

Model	Accuracy(%)		
	positive	negative	overall
RNTN	69.83	70.17	69.93
SVM1			
SVM2			

Table 5: Experiment 2

- Exp 3: Training on *Sentiment Treebank*, label the training set of *SemEval*, retrain the model with both *Sentiment Treebank* and the testset of *SemEval*, and test the new model on the test set of *SemEval*.

Model	Accuracy(%)		
	positive	negative	overall
RNTN			
SVM1			
SVM2			

Table 6: Experiment 3

7 Conclusion

Better parser needed.

Acknowledgments

We would like to thanks Dr. Mooney for his guidance. Thanks researcher who share their tweet corpus.

References

[Aho and Ullman1972] Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

[American Psychological Association1983] American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

[Association for Computing Machinery1983] Association for Computing Machinery. 1983. *Computing Reviews*, 24(11):503–512.

[Chandra et al.1981] Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

[Gusfield1997] Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.