CMPE 561 - Final Presentation

NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets

D. Kekulluoglu, G.E. Kaya, G.Yesiltas

Outline

- Introduction
- Approach of the Selected Paper
- Our Approach
- Results of Experiments
- Conclusion

Introduction

- Given a tweet, decide whether the message is of positive, negative, or neutral sentiment.
- SVM classifier to detect sentiment.

Approach of the Selected Paper

- Each tweet represented as a feature vector made up of:
  - Word n-grams
  - Character n-grams
  - All caps
  - POS
  - Hashtags
  - Lexicons
  - Punctuation
  - Emoticons
  - Elongated words
  - Clusters
  - Negation

Results of the Selected Paper

- Trained on set of 9912 annotated tweets.
- Tested on 3813 unseen tweets.

	Classifier	Tweets	SMS
Training set:	Majority	26.94	-
	SVM-all	67.20	-
Development set:	Majority	26.85	-
	SVM-all	68.72	-
Test set:	Majority	29.19	19.03
	SVM-unigrams	39.61	39.29
	SVM-all	**69.02**	**68.46**

Our Approach

- Features Used
  - Word n-grams
  - POS
  - Lexicons
  - Emoticons
  - Negation
- For every training tweet extract feature vectors then train the SVM.
- For every test tweet extract feature vectors then predict using SVM.
- In total 5422 features used for the biggest feature vector. Really sparse

Dataset

- 5345 training tweets, 1506 test tweets
- In the training set;
    - 1840 positive,
    - 2752 neutral,
    - 753 negative.

POS

- Tokens created using CMU Pos Tagger was given with dataset.
- There are 25 tags.
- Twitter/online-specific tags:
  - #: hashtag
  - @: at-mention
  - ~: discourse marker, indications of continuation of a message across multiple tweets
  - U: URL or email address
  - E: emoticon
- The number of occurrences of each part-of-speech tag is used as feature.

Negation

- The number of negated context is used as feature.
- Negated context is defined as:
  - Starts with a negation word (such as ***no, don't, shouldn't***)
  - Ends with one of the punctuation marks: '***,***', '***.***', '***:***', '***;***', '***!***', '***?***'
- The list of negation words was adopted from Christopher Potts' sentiment tutorial.
- Negated context also affects n-grams and lexicon feature.

Word N-grams

- Tokenize training tweets (Outputs of CMU Pos Tagger is used)
- Create a dictionary with all unary, binary and ternary words
- Prune the features by taking;
  - Unary words with at least 3 occurrences.
  - Binary words with at least 5 occurrences.
  - Ternary words with at least 7 occurrences.
- Extract features from tweets using these dictionaries.
- We have 5377 features combining these unary,binary and ternary words.
- For every feature, we look at the tweet;
  - If that feature is in the tweet write the number of occurrences. I.e.“So so so beautiful” has 3 for ‘so’ feature.
  - Otherwise write 0.
- Very sparse feature vectors.

Adjectives

- The idea is instead of using all words for feature vector, use adjectives since they are more descriptive.
- Only unigrams are created.
- Same principles as before.
- 90% decrease in the feature vector size.
- Better performance when we took default C for SVM.
- However no significant change when we take the C as 0.005.

Our Approach

- Lexicons Used;
  - NRC Emotion Lexicon
  - MPQA Lexicon
  - Bing Liu Lexicon
  - NRC Hashtag Sentiment Lexicon
  - Sentiment140 Lexicon
  - DGE Emoticon Lexicon :)

Lexicons

- NRC Emotion Lexicon
  - 14,182 words,
  - either negative or positive,
  - Score value 1 or -1.
- MPQA Lexicon
  - 8222 words,
  - either negative or positive,
  - Score value 1 or -1.

Lexicons

- Bing Liu Lexicon
  - 4783 negative, 2006 positive words,
  - Score value 1 or -1.
- NRC Hashtag Sentiment Lexicon
  - 54,129 words, consisting of mostly hashtags.
  - Score value is between -5 and 5.

Lexicons

- Sentiment140 Lexicon
  - 62,468 words, including hashtags, usernames and emoticons
  - Score value is between -5 and 5.
- DGE Emoticon Lexicon :)
  - Manually annotated
  - Calculated considering the number of occurrences of each emoticon in each class.

Lexicons

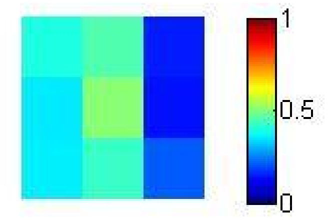
- 4 features extracted from lexicons;
  - total count of tokens in the tweet with score(w, p) > 0;
  - total score of the tweet ;
  - the maximal word score in tweet ;
  - the score of the last token in the tweet with score(w, p) > 0;

Classification

- SVM
- 3 classes; Positive, Negative and Neutral
- C = 0.005 or C = 1
- Extra penalty weight for misclassification for negative tweets - because underrepresented
- Adjectives only features
- DGE Emoticon lexicon

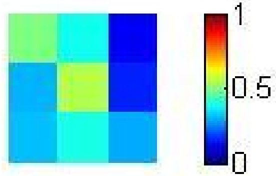
Result #1

- All word n-grams + POS + SentimentLexicon140 + Negation
- C = 1, no penalty
- Accuracy = 41.10%
- F1 measure = **<u>38.32%</u>**



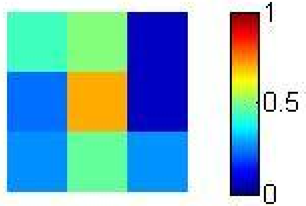
Result #2

- All word n-grams + POS + All lexicons (except DGE EmoLex) + Negation
- C = 1, no penalty
- Accuracy = 47.74%
- F1 measure = **<u>44.81%</u>**



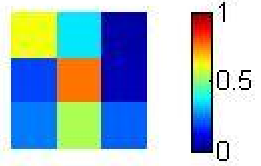
Result #3

- Adjectives unigrams + POS + All lexicons (except DGE EmoLex) + Negation
- C = 1, no penalty
- Accuracy = 52.12%
- F1 measure = **<u>49.14%</u>**



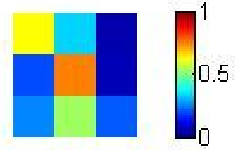
Result #4

- All word n-grams + POS + All lexicons (except DGE EmoLex) + Negation
- C = 0.005, no penalty
- Accuracy = 59.56%
- F1 measure = **<u>55.33%</u>**



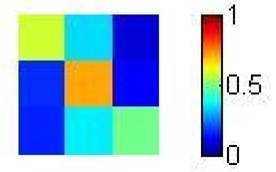
Result #5

- Adjective unigrams + POS + All lexicons (except DGE EmoLex) + Negation
- C = 0.005, no penalty
- Accuracy = 59.56%
- F1 measure = **55.16%**



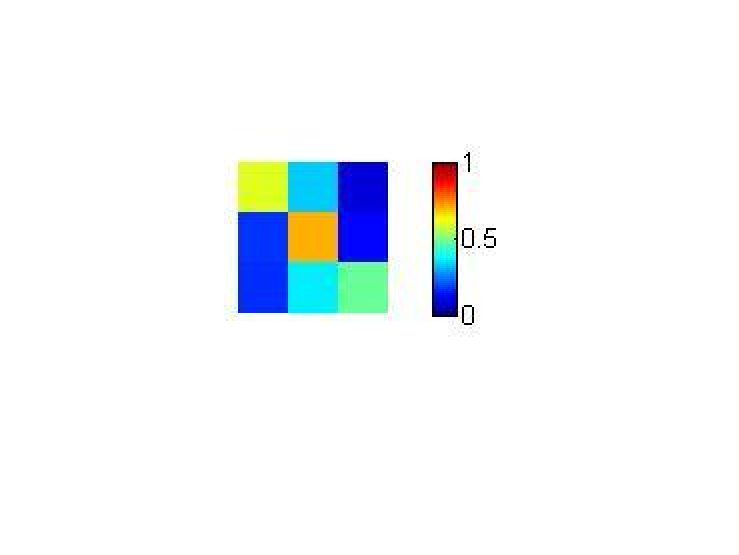
Result #6

- All word n-grams + POS + All lexicons (except DGE EmoLex) + Negation
- C = 0.005, penalty for negative
- Accuracy = 61.82%
- F1 measure = **<u>59.89%</u>**



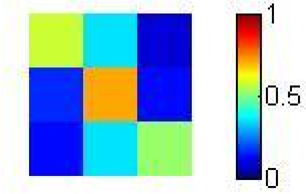
Result #7

- Adjective unigrams + POS + All lexicons (except DGE EmoLex) + Negation
- C = 0.005, penalty for negative =
- Accuracy = 61.68%
- F1 measure = **<u>59.58%</u>**



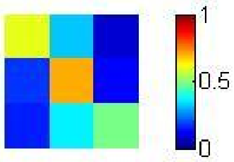
Result #8

- All word n-grams + POS + All lexicons (with DGE EmoLex) + Negation
- C = 0.005, penalty for negative
- Accuracy = 62.28%
- F1 measure = **60.72%**



Result #9

- Adjective unigrams + POS + All lexicons (with DGE EmoLex) + Negation
- C = 0.005, penalty for negative =
- Accuracy = 62.48%
- F1 measure = **60.57%**



All Results

	Accuracy	**F-score**
Result #1	41.10	38.02
Result #2	47.74	44.81
Result #3	52.12	49.14
Result #4	59.56	55.33
Result #5	59.56	55.16
Result #6	61.82	59.89
Result #7	61.68	59.58
Result #8	62.28	**60.72**
Result #9	62.48	60.57

References

1. Mohammad, Saif M., Svetlana Kiritchenko, and Xiaodan Zhu. "NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets." *arXiv preprint arXiv:1308.6242* (2013).
2. Kökciyan, Nadin, et al. "Bounce: Sentiment classification in Twitter using rich feature sets." *Second Joint Conference on Lexical and Computational Semantics (* SEM)*. Vol. 2. 2013.
3. Mohammad, Saif M., and Peter D. Turney. "Crowdsourcing a word–emotion association lexicon." *Computational Intelligence* 29.3 (2013): 436-465.
4. Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005.
5. Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews."*Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.
6. Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
7. Gimpel, Kevin, et al. "Part-of-speech tagging for twitter: Annotation, features, and experiments." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 2011.

Thanks for your attention

Any question?