SeNTU: Sentiment Analysis of Tweets by Combining a Rule-based Classifier with Supervised Learning

Prerna Chikersal, Soujanya Poria, and Erik Cambria
School of Computer Engineering
Nanyang Technological University
Singapore - 639798
{prerna1,sporia,cambria}@ntu.edu.sg

Abstract

We describe a Twitter sentiment analysis system developed by combining a rule-based classifier with supervised learning. We submitted our results for the message-level subtask in SemEval 2015 Task 10, and achieved a F[1]-score of 57.06%. The rule-based classifier is based on rules that are dependent on the occurrences of emoticons and opinion words in tweets. Whereas, the Support Vector Machine (SVM) is trained on semantic, dependency, and sentiment lexicon based features. The tweets are classified as *positive*, *negative* or *unknown* by the rule-based classifier, and as *positive*, *negative* or *neutral* by the SVM. The results we obtained show that rules can help refine the SVM's predictions.

1 Introduction

Our opinions and the opinions of others play a very important role in our decision-making process and even influence our behaviour. In recent times, an increasing number of people have taken to expressing their opinions on a wide variety of topics on microblogging websites such as Twitter. Being able to analyse this data and extract opinions about a number of topics, can help us make informed choices and predictions regarding those topics. Due to this, sentiment analysis of tweets is gaining importance across a number of domains such as e-commerce (Wang and Cardie, 2014), politics (Tumasjan et al., 2010; Johnson et al., 2012; Wang et al., 2012), health and psychology (Cambria et al., 2010; Harman, ; Harman,), multimodality (Poria et al., 2015), crowd validation (Cambria et al., 2010), and even intelligence and surveillance (Jansen et al., 2009).

SemEval 2015 Task 10 (Rosenthal et al., 2015) is an international shared-task competition that aims to promote research in sentiment analysis of tweets by providing annotated tweets for training, development and testing. We created a sentiment analysis system to participate in the message-level task of this competition. The objective of the system is to label the sentiment of each tweet as "positive", "negative" or "neutral".

In this paper, we describe our sentiment analysis system, which is a combined classifier created by integrating a rule-based classification layer with a support vector machine.

2 System Description

Our Sentiment Analysis System consists of two classifiers – (i) Rule-based and (ii) Supervised, integrated together. This section describes both these classifiers and how we combine them.

During pre-processing, all the @<*username*> references are changes to @*USER* and all the *URLs* are changed to *http://URL.com*. Then, we use the CMU Twitter Tokeniser and POS Tagger (Gimpel et al., 2011) to tokenise the tweets and give a parts-of-speech tag to each token. We use the POS tags to remove all emoticons from the pre-processed tweets. Pre-processed tweets **with emoticons** are given as input to the rule-based classifier, whereas the support vector machine takes pre-

[1]We average the positive and negative F-measures to get the F-score, which is the evaluation metric for this task.

Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 647–651,
Denver, Colorado, June 4-5, 2015.

processed tweets **without emoticons** as an input.

2.1 Supervised Learning

For the supervised classifier, we cast the sentiment analysis problem as a multi-class classification problem, where each tweet has to be labeled as "positive", "negative" or "neutral". We train a Support Vector Machine (SVM) (Cortes and Vapnik, 1995) on the tweets provided for training. For all our experiments, we use a linear kernel and L1-regularisation. The C parameter is chosen by cross-validation. As mentioned above, emoticons have already been removed from tweets given as input to the SVM.

Each tweet is represented as a feature vector, containing the following features:

- **Word N-grams:** Frequencies of contiguous sequences of 1, 2 or 3 tokens. The TF-IDF weighting scheme is applied.

- **Character N-grams:** Frequencies of contiguous sequences of 1, 2 or 3 characters inside each word's boundary. The TF-IDF weighting scheme is applied.

- **POS Tags:** Using CMU Twitter Tagger (Gimpel et al., 2011) output, for each tweet we compute – (i) *countAdj* (number of adjectives), (ii) *countAdv* (number of adverbs), (iii) *countNoun* (number of nouns, proper nouns, and proper nouns+possessives), (iv) *countVerb* (number of verbs), and (v) *countIntj* (number of interjections). The sum of these five counts, gives us the *totalPos*. The POS features are: $[\frac{countAdj}{totalPos}, \frac{countAdv}{totalPos}, \frac{countNoun}{totalPos}, \frac{countVerb}{totalPos}, \frac{countIntj}{totalPos}]$.

- **@USER:** A boolean feature that is set to 1 if the tweet contains a @<*username*> reference.

- **Hashtag:** A boolean feature that is set to 1 if the tweet contains a hashtag.

- **URL:** A boolean feature that is set to 1 if the tweet contains a URL.

- **Discourse:** A boolean feature that is set to 1 if the tweet contains a "discourse marker". Examples of discourse markers would be a "RT" followed by a username to indicate that the tweet is a re-tweet, news article headline followed by "..." followed by a URL to the news article, etc. Basically, this feature indicates whether or not the tweet is a part of a discourse.

- **Sentiment140 Lexicon:** The Sentiment140 Lexicon (Mohammad et al., 2013) contains unigrams and bigrams along with their polarity scores in the range of -5.00 to $+5.00$. Considering all uni/bi-grams with polarity less than -1.0 to be negative and with polarity greater than $+1.0$ to be positive, we count the number of negative (*negativesCount*) and the number of positive (*positivesCount*) uni/bi-gram occurrences in every tweet. For each tweet,
  - the *polarityMeasure* is based on the *positivesCount* and *negativesCount*, and calculated using Algorithm 1.
  - the maximum polarity value (*maxPolarityValue*) is the most positive or most negative polarity value of all polar uni/bi-gram occurrences in the tweet.

  Both these features are normalised to values between -1 and $+1$.

Algorithm 1 Calculating *polarityMeasure* based on *positivesCount* and *negativesCount*

if positivesCount $>$ negativesCount **then**
    if negativesCount $!= 0$ **then**
        polarityMeasure $= \frac{positivesCount}{negativesCount}$
    else
        polarityMeasure $=$ positivesCount
    end if
else if negativesCount $>$ positivesCount **then**
    if positivesCount $!= 0$ **then**
        polarityMeasure $= -1 \times \frac{negativesCount}{positivesCount}$
    else
        polarityMeasure $= -1\times$ negativesCount
    end if
end if

- **Bing Liu Lexicon:** The Bing Liu lexicon (Liu et al., 2005) is a list of positive and negative words. We count the number of positive (*positivesCount*) and negative words (*negativesCount*) in each tweet, and calculate *polarityMeasure* using Algorithm 1. The *polarityMeasure* is appended to the feature vector.

- **NRC Emotion Lexicon:** The NRC Emotion Lexicon (Mohammad and Turney, 2013) contains a list of positive and negative words. The *polarityMeasure* is calculated using the method used for the Bing Liu Lexicon.

- **NRC Hashtag Lexicon:** The NRC Hashtag Lexicon (Mohammad et al., 2013) contains unigrams and bigrams along with their polarity scores in the range of -5.00 to $+5.00$. Using the method used for the Sentiment140 Lexicon, we calculate *polarityMeasure* and *maxPolarityValue*, and append them to the feature vector.

- **SentiWordNet:** SentiWordNet (Esuli and Sebastiani, 2006) assigns to each synset of WordNet (Fellbaum, 2010) 3 scores: positivity, negativity, objectivity. A word whose positivity score is greater than negativity and objectivity is positive, while a word whose negativity score is greater than positivity and objectivity is negative. For each tweet, we calculate *polarityMeasure* and *maxPolarityValue* using the method used for the Bing Liu Lexicon.

- **SenticNet:** SenticNet (Cambria et al., 2014) contains polarity scores of single and multiword phrases. We count the number of positive and negative words/phrases in each tweet, and calculate *polarityMeasure* using the method used for the Sentiment140 Lexicon.

- **Negation:** The Stanford Dependency Parser (De Marneffe et al., 2006) is used to find negation in tweets. Negation is not a feature on its own. Rather, it affects the word n-grams and the lexicons related features. The negated word is appended with a "_NEG" in all n-grams, while the polarity of all negated words is inverted in the lexicon features.

2.2 Rule-based Classifier

For the rule-based classifier, we cast the problem as a multi-class classification problem, where each tweet is to be labeled as "positive", "negative" or "unknown". This is an unsupervised classifier, which applies the following rules for predictions:

- **Emoticon-related Rules:** If a tweet contains only positive emoticons and no negative emoticons, it is classified as positive. If a tweet contains only negative emoticons and no positive emoticons, it is classified as negative. If a tweet contains no emoticons, we apply the sentiment lexicon-related rules. The following emoticons are considered to be positive: :) , (: , ;) , :-) , (-: , :D , :-D , :P , :-P . While, the following emoticons are considered to be negative: :(,): , ;(, :-(,)-: , D: , D-: , :'(, :'-(,)': ,)-': .

- **Sentiment Lexicon-related Rules:** The Bing Liu lexicon, the NRC Emotion lexicon, and SentiWordNet are used as resources for positive and negative opinion words. If a tweet contains **more than two** positive words, and no negation or negative words from either of the lexicons, it is classified as positive. If a tweet contains **more than two** negative words, and no negation or positive words from either of the lexicons, it is classified as negative. If none of the above rules apply, the tweet is classified as unknown.

2.3 Combining the Classifiers

After developing the rule-based classifier and training the SVM, we combine the them to refine the SVM's predictions. Since, our goal is to maximise positive and negative precision and recall, we use the rule-based classifier to correct or verify the "neutral" SVM predictions. So, for every tweet labeled as neutral by the SVM, we consider the predictions of the rule-based layer as the final labels.

3 Experiments and Results

We trained a Support Vector Machine (SVM) on 9418 tweets allowed to be used for training purposes. The results we submitted to SemEval 2015 were yielded by using all SVM features and emoticon-related rules. The sentiment lexicon-related rules were implemented later, and thus could not be used for the official submission. Table 2 shows the official test results for SemEval 2015.

Features	Positive			Negative			Neutral			F_{pn}
	P	R	F	P	R	F	P	R	F	
All Features	0.824	0.629	0.713	0.612	0.607	0.610	0.679	0.831	0.748	0.662
w/o N-grams	0.671	0.597	0.632	0.430	0.574	0.491	0.645	0.637	0.641	0.562
w/o POS Tags	0.814	0.611	0.698	0.633	0.589	0.610	0.669	0.839	0.744	0.654
w/o @User, Hashtag, URL, Discourse	0.821	0.616	0.704	0.602	0.607	0.605	0.672	0.826	0.741	0.654
w/o Sentiment140	0.814	0.616	0.701	0.602	0.599	0.600	0.676	0.830	0.745	0.651
w/o Bing Liu	0.821	0.621	0.707	0.616	0.603	0.610	0.676	0.833	0.746	0.658
w/o NRC Emotion + Hashtag	0.816	0.619	0.705	0.609	0.597	0.603	0.676	0.832	0.746	0.654
w/o SentiWordNet	0.821	0.624	0.709	0.610	0.597	0.603	0.674	0.830	0.744	0.656
w/o SenticNet	0.820	0.615	0.703	0.610	0.597	0.603	0.674	0.837	0.747	0.653
w/o Negation	0.811	0.610	0.701	0.598	0.601	0.593	0.674	0.824	0.744	0.647

Table 1: Feature ablation study for the SVM classifier. Each row shows the precision, recall, and F-score for the positive, negative, and neutral classes respectively, followed by the average positive and negative F-score, which is the chosen evaluation metric. All values in the table are between 0 and 1, and are rounded off to 3 decimal places.

Dataset	Our Score	Best Score
Twitter 2015	57.06	64.84
LiveJournal 2014	68.70	75.34
Twitter 2014	66.85	74.42
Twitter 2013	63.50	72.80
SMS 2013	60.53	68.49
Twitter 2014 Sarcasm	45.18	57.50

Table 2: Average positive and negative F-scores for system with all SVM features and only emoticon rules.

Table 1 reports the results of a feature ablation study carried out by testing the SVM classifier on 3204 development tweets (from SemEval 2013) not included in the training data. These are cross-validation results obtained using the hold-out method.This study helps us understand the importance of different features. From the table, we can see that the word and character n-grams features are the most useful, followed by negation and then the rest. All sentiment lexicon related features appear to have similar importance, but we get the best F-score when we append them all to the feature vector.

Features	F_{pn}	Classification Rate (%)
All Features	66.2	71.5
All Features and Rules	66.7	72.3

Table 3: Comparison between the results obtained using SVM alone, and using SVM with a rule-based layer.

Since, using all the previously described features gives the best SVM predictions, we add the rule-based classification layer to a SVM trained on all features. Table 3 compares the results obtained using the SVM alone with the results obtained using SVM along with all the rules (emoticon and lexicon-based) specified in section 2.2. We observe that the F-score further increases by around half a unit and the classification rate[2] increases by around 0.8.

4 Conclusion

In this paper, we described a sentiment analysis system developed by combining a SVM with a rule-based classification layer. Even though we do not get the best scores, we find that a rule-based classification layer can indeed refine the SVM's predictions. We also devise creative twitter-specific, negation and lexicon-related features for the SVM, and demonstrate how they improve the sentiment analysis system. In future, we aim to use enriched sentiment and emotion lists like the ones used by (Poria et al., 2012). We would also like to experiment with refining the SVM's predictions using more rules based on complex semantics.

Acknowledgments

We wish to acknowledge the funding for this project from Nanyang Technological University under the Undergraduate Research Experience on CAmpus (URECA) programme.

[2]Classification rate $= \frac{number\ of\ tweets\ classified\ correctly}{total\ number\ of\ tweets}$

References

Erik Cambria, Amir Hussain, Catherine Havasi, Chris Eckl, and James Munro. 2010. Towards crowd validation of the uk national health service. *WebSci10*, pages 1–5.

Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. 2014. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Twenty-eighth AAAI conference on artificial intelligence*, pages 1515–1521.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454.

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422.

Christiane Fellbaum. 2010. Wordnet: An electronic lexical database. 1998. *WordNet is available from http://www. cogsci. princeton. edu/wn*.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 42–47.

Glen Coppersmith Mark Dredze Craig Harman. Quantifying mental health signals in twitter.

Bernard J Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. 2009. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188.

Christopher Johnson, Parul Shukla, and Shilpa Shukla. 2012. On classifying the political sentiment of tweets. *cs. utexas. edu*.

Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342–351.

Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Saif Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA, June.

Soujanya Poria, Alexander Gelbukh, Erik Cambria, Peipei Yang, Amir Hussain, and Tariq Durrani. 2012. Merging senticnet and wordnet-affect emotion lists for sentiment analysis. In *Signal Processing (ICSP), 2012 IEEE 11th International Conference on*, volume 2, pages 1251–1255.

Soujanya Poria, Erik Cambria, Amir Hussain, and Guang-Bin Huang. 2015. Towards an intelligent framework for multimodal affective data analysis. *Neural Networks*, 63:104–116.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, SemEval '2015, Denver, Colorado, June.

Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10:178–185.

Lu Wang and Claire Cardie. 2014. A piece of my mind: A sentiment analysis approach for online dispute detection. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, volume 2, pages 693–699.

Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, pages 115–120.