Crowd Sentiment Detection during Disasters and Crises

Ahmed Nagy
Carnegie Mellon Silicon Valley
IMT Lucca Institute of Advanced Studies
ahmed.nagy@gmail.com

Jeannie Stamberger
Disaster Management Initiative
Carnegie Mellon Silicon Valley
jeannie.stamberger@gmail.com

ABSTRACT

Microblogs are an opportunity for scavenging critical information such as sentiments. This information can be used to detect rapidly the sentiment of the crowd towards crises or disasters. It can be used as an effective tool to inform humanitarian efforts, and improve the ways in which informative messages are crafted for the crowd regarding an event. Unique characteristics of microblogs (lack of context, use of jargon etc) in Tweets expressed by a message-sharing social network during a disaster response require special handling to identify sentiment. We present a systematic evaluation of approaches to accurately and precisely identify sentiment in these Tweets. This paper describes sentiment detection expressed in 3698 Tweets, collected during the September 2010, San Bruno, California gas explosion and resulting fires. The data collected was manually coded to benchmark our techniques. We start by using a library of words with annotated sentiment, *SentiWordNet 3.0*, to detect the basic sentiment of each Tweet. We complemented that technique by adding a comprehensive list of emoticons, a sentiment based dictionary and a list of out-of-vocabulary words that are popular in brief, online text communications such as *lol*, *wow*, etc. Our technique performed 27% better than Bayesian Networks alone, and the combination of Bayesian networks with annotated lists provided marginal improvements in sentiment detection than various combinations of lists.

Keywords
Crisis management, disaster response, emergency management, sentiment detection, Twitter, short messages

I. INTRODUCTION

Our work addresses the sentiment expression during disaster response that is exchanged using platforms that produce short messages such as Tweets and SMS. Social media platforms such as Twitter produce fast-spreading messages generated by a geographically dispersed crowd. Having an effective fast method to detect sentiment can help respond effectively to the sentiment state of the crowd. Further it can be used to craft strategies for information dissemination to the crowd. In 2010, a gas pipeline explosion occurred at 6:11 p.m. PDT on September 9, 2010, in San Bruno, California, a suburb of San Francisco, when a 30 inch diameter steel natural gas pipeline owned by Pacific Gas & Electric exploded into flames 2 miles west of San Francisco International Airport. Immediately afterwards social media platforms including Twitter started exchanging messages regarding the event. We collected 3698 Tweets about the event during the first 24 hours. Our aim is to process these message streams to understand better how people feel, and to respond more appropriately, perhaps through official messaging.

Tweet	Sentiment Value
#SanBrunoFire What law are they using to keep the press away? Obviously it is not all a crime scene? #NannyState officials making up law.	Negative
Thank you to all the brave firefighters helping to save lives in the #SanBrunoFire and #BoulderFire http://bit.ly/tyfirefighters	Positive

Table 1. Tweet Examples with Sentiment Value

One of the important approaches to respond to the situation is to detect the absence of explicit sentiment information, for example, Table 1 shows negative sentiment information in the first Tweet. The two Tweets are extracted from the collected San Bruno fire Tweets. The second Tweet is considered with a positive sentiment. Even though it states that there are victims, it salutes the efforts of firefighters to contain the situation. As a result, the sentiment information can be used to project the information to the crowd in better ways. In addition, calls to action can be requested that can improve the situation, for example calls to blood donations or money donations. This can be discovered from the sentiment in the Tweets. For example, sentiment detection and analysis can be used for a relevant corpus of Tweets to give an estimation of the level of devastation or situation recovery. As a result, the sentiment analysis can be further extended to infer the effectiveness of the existing response in minimizing the emotional impact of a disaster.

Our sentiment detection framework is effective at detecting the sentiment of the crowd; it can act as a thermometer to measure sentiment towards an event. Further, it can be used to measure the sentiment towards a message that is received through social media. We have been working with Cupertino county to craft effective methods to reach users through social networks. Using this framework is an effective means to give a quick measure to the impact of a message through social networks. We are aware of the limitation of relying solely on the sentiment collected from the social networks to be presented as the sentiment of the crowd. However, it can still give a feeling or a pulse of the crowd towards an event. This situation can be explained further by looking deeper into the type of users interested to share their feelings on social networks and the representation of the sample. It is also a challenge to guarantee a representative sample of a set of users experiencing an event and sharing their feelings on a social network.

One of the fundamental conclusions we reached from the first 24-hours of #SanBrunoFire Tweets is that most of the Tweeters did not receive information about the event through first-hand knowledge sources. It can be observed from the data gathered that several Tweets relied on radio, television, and online news sources reported disaster response status many hours earlier than official sources. It might be useful to separate the data collected from the geographical area facing the disaster or the crisis from the rest of the data. This can help improve the accuracy of determining the sentiment of a crowd facing an emergency. Furthermore, analyzing the message content can give an estimate about the effectiveness of the media methods to reach the crowd either in the disaster areas or the other areas.

Short messages spread by tools such as Twitter lack enough context. Short messages are likely to contain emoticons, links, out of vocabulary words and swear words. Several of the techniques developed for assessing sentiment for long passages of text might not fully fit the constraints inherent in short text messages. For the data collected, only 34 posts had emoticons. Several approaches have been used to evaluate the level of sentiment for short messages. The techniques can be summarized as using classifiers, dictionary lists or a combination. One of our contributions is developing a technique that is generic enough to detect and analyze sentiment expressed in short messages. In addition, apply the technique on data that is relevant to a disaster event. A generic framework works with minimal tweaking and customization to detect sentiment for new events and disasters, which results in saving time and effort. Another contribution is the development of a methodology that can be used to collect and analyze the emotional pulse of the crowd automatically during a disaster.

Here we explore a variety of methods to identify sentiment in disaster Tweets to test the following hypotheses: 1)The sentiment of Tweets changes over time in disaster recovery (time since the disaster) and 2) Information-based Tweets have more nouns than adjectives with and low sentiment content.
This paper is organized as follows: section II presents relevant prior work; section III describes the methods, including a description of the system architecture, followed by a discussion on the system limitation; section IV describes the results; and, section VI closes with the conclusion and further research.

II. RELEVANT PRIOR WORK

Sentiment analysis and discovery has acquired a lot of importance recently. Sentiment detection from microblogs can yield a fast way to get a heartbeat of the crowd. Here we highlight relevant research in the area of sentiment detection in Tweets. Sentiment detection methods include two main approaches: (1) annotated word lists and (2) classifiers, where the classifier depends on the co-occurrence of sentiment words to cluster the text with sentiment; some researchers have employed a hybrid of these two approaches. The annotated word list approach consists of words in which sentiment value has been pre-identified. In this approach, word list expansion is sometimes used to grow the initial seed of annotated sentiment words, using techniques such as clustering similar words and looking up dictionaries for synonyms.

Nielsen (2011) presented a list-based approach to calculate the Tweet sentiment. Nielsen used an initial list from a dictionary. He expanded the lists by manually examining Tweets that scored high for sentiment values and expanded the set of sentiment lexicons by including the antonyms and synonyms of such lexica. The list had values ranging from +5 to -5 for sentiment with 2477 unique words. Surprisingly, Nielsen's simple matching approach performed better than the more comprehensive list, Affective Norms for English Words (ANEW) presented by Bradley and Lang (1999).

Two classification approaches identified different levels of sentiment in Tweets. Agarwal, Xie, Vovsha, Rambow, and Passonneau (2011) presented a classification-based approach to detect sentiments in Tweets. Manually-annotated Tweets were used to extract sentiment values. The supervised models presented resulted in better sentiment detection than the gold standard examined and more compact number of features. Decision trees were used as a way to classify the new Tweets. A sentiment dictionary was used to calculate the sentiment polarity of words. It contained 8000 words with scores between 1 - 3 where 1 is negative and 3 is positive. Two models were investigated with the "tree kernel": bigram and unigram model. The unigram model outperformed the bigram model which included parts of speech tags. Adding other features such as emoticons, word polarity, and exclamation marks were used to enhance the unigram model. Compared against the gold standard, the tree kernel model achieved an accuracy of 60 % compared to 56 % for the feature-based approach. Barbosa and Feng (2010) introduced a classification method where they concluded (based on the training of data) that the top five features in terms of information gain are: negative polarity, positive polarity, verbs, emoticons and upper case. They applied a summation and normalization approach to calculate the sentiment of Tweets gathered from twendz, Twittersentiment and Tweetfeel. Interestingly, they found that four positive words (*awesome, rock, love* and *beat*) express 95 % of the positive sentiment while 96 % of the negative sentiment is expressed by six words (*hate*, *suck*, *wtf*, *piss*, *stupid* and *fail*).

Several methods include social network characteristics to detect sentiment, such as influence of the user, sentiment of user's followers, or a sentiment profile of a user. Hui and Gregory (2010) created a light form of formalization for quantifying sentiment and influence in microblogs, especially Twitter. The sentiment of a Tweet or a post was calculated as the mean of the sentiment values in a comment. The algorithm takes into consideration the influence of the user who posted the Tweet. The influence was computed as a function of the total number of comments, relevancy to a topic and total number of followers. The centrality measure can be misleading. Since followers might have different opinions on the same topics, it might not be a valid reason to propagate a sentimental state among all followers. In fact, this statement should be taken cautiously since sometimes the user changes his sentiment towards an event several times in a very short time, such as during sports matches where Tweets supporting and criticizing a team are posted simultaneously! Similarly, Tan Lee, Tang, Jiang, Zhou, and Li (2011)) presented an approach where they took into consideration the sentiment of a users social network to calculate his/her overall sentiment. The main assumption is that connected users are more likely to hold similar opinions. Incorporating this assumption into support vector machines revealed greater degree of sentiment inference when a user's social network is taken into consideration. The experiments were based on collecting the sentiment of followers of networks of followers of politicians (e.g.,Obama), where followers and followee may tend to have more similar sentiments than networks based on non political interests. Pennacchiotti and Popescu (2011) presented a classification approach based on trying to build a profile for users from the topics they post. The profile is further analyzed to mine for the sentiment. They included the effect of the social network on the sentiment of a user. They calculate the sentiment of a user towards a specific topic as the absolute difference between positive and negative sentiment normalized by the total number of words in a Tweet. The assumptions are very similar to those in Tan, Lee, (Tang, Jiang, Zhou, and Li 2011). The features selected for the classification algorithm are based on frequency. As a result, the most frequent words are used to classify the users and build their profiles. The experiments were also run on a set for users interested in political posts.

Building on this prior work, we explore the precision and accuracy of several methods such as Bayesian networks, sentiment lists and sentiment based ontology to identify sentiment of Tweets in the particular use case of heightened emotions during disasters.

III. METHODS

We aimed at developing methods that can accurately detect the sentiment value of Tweets in a disaster context. We integrated several methods to ensure the flexibility of the techniques. We compared 4 different methods which included different combinations of the approaches described below: SentiWordNet, Emoticons, AFNN and Bayesian Networks. The 4 methods we compared included the following combinations of approaches: 1)

SentiWordNet, Emoticons and AFNN 2) SenitWordNet and Emoticons 3) Bayesian Network 4) SentiWordNet, Emoticons, AFNN, and Bayesian Network.

WordNet

We started by considering WordNet (Miller 1995) as an ontology to give polarity for the words used. However, we soon discovered that the relatedness among words is not symmetric. Table 2 illustrates an example of asymmetry for disaster-related terms. Words that may be considered related in disasters and disaster jargon (e.g., disaster and earthquake) show little WordNet relatedness: while the word relatedness of "Problem" and "Disaster" is 0.6 and between "Earthquake" and "Problem" is 0.5, the relatedness value between "Disaster" and "Earthquake" is a magnitude lower than 0.05! Therefore, for domain-specific use cases such as disaster response, as the focus is here, WordNet relatedness values may not be appropriate.

Word A	**Word B**	**WordNet relatedness between words A and B**
Problem	Disaster	0.60
Disaster	Earthquake	0.05
Earthquake	Problem	0.50

Table 2. WordNet Relatedness Value Mismatch for Disaster-related Terms

SentiWordNet

We used SentiWordNet (Baccianella, Esuli, and Sebastiani 2008) to detect the sentiment value of Tweets. SentiWordNet is a lexical resource explicitly devised for supporting sentiment classification and opinion mining applications. Each word (or synset, in WordNet terminology) included in SentiWordNet has three characteristics: positivity, negativity and neutrality. To calculate the sentiment orientation of the word according to Baccianella, Esuli, and Sebastiani (2008) the following equation is used

SentimentOrientation= 1 – (PositiveScore - NegativeScore).

SentiWordNet is based on first giving the lexicon sentiment values by a human being then using this as a seed to establish the values for other lexicons. Support vector machines were used to classify the lexicons linked to the seed. SentiWordNet is considered a general method to detect the sentiment in a text. We found out that some words have sentiment values even though they should be considered neutral e.g., here, a, sentiment, exist. We developed a list of words that should not be considered in evaluating the sentiment value of the Tweet to avoid accumulating sentiment value for words that should be considered neutral.

Emoticons

Emoticons are a popular way that users express their emotions towards a topic or an entity. Unlike words, emoticons are small pictures or expressions created with symbols (e.g., the popular smiley face ":)" , which has a positive sentiment value). We incorporated a comprehensive list of emoticons annotated with sentiment values (SentiStrength, 2011). The list has 169 different emoticons having sentiment value ranging from -2 to +2. Emoticons are usually widespread in social network postings, as a result, it is important to develop a technique for detecting and scoring them sentimentally.

AFNN

We incorporated the sentiment list AFNN, presented in (Nielsen, 2011). It has 2477 words with values ranging from -3 to +3. This is a comprehensive list of the most frequent lexicons that carry sentiment values.

Sentiment Value Calculation

We experimented with two methods to calculate the sentiment value of a Tweet: (1) sentiment value count and (2) normalized sentiment value count. In the first method, we sum the sentiment value of words identified as having sentiment. The following equation is used to calculate the sentiment value as an integer:

$$Sentiment\ Value = \sum_{i=0}^{n} (P_i - N_i)$$

where

P_i: Number of words with positive sentiment value.

N_i: Number of words with negative sentiment value.

n: total number of lexicons in a Tweet.

In the second method, we normalize the sum of the sentiment value by multiplying by the ratio of the total number of sentiment words and total number of words in a Tweet. The following equation is used to calculate the sentiment orientation of a Tweet. The value is a normalized number between -1 and 1 showing the sentiment orientation of the Tweet

$$Normalised\ Sentiment\ Value = \frac{\sum_{i=0}^{n}(P_i - N_i)}{Total\,Number\,of\,Words}$$

This normalization technique is borrowed from the inverse document frequency method (Yu and Mizuno, 1988). However, in microblogs, there is little opportunity for words to repeat compared with long passages. As a result, two messages one short and the other long having all their words with sentiment value will have the same sentiment strength which might be misleading.

Naive Bayesian Networks

We used a Naive Bayesian classifier to feed some of the Tweets or sentences that are already classified then we used it to find the sentiment value for new Tweets. Preliminary results show that it is not enough to train the Bayesian classifier with instances that are manually classified. Incorporating emoticons, SentiWordNet and AFNN increased the accuracy of classification by approximately 18 %. However, incorporating AFNN did not increase it significantly, only an improvement of 9 % for the calculation of the sentiment value for the data set run. When we investigated the reasons behind that we discovered that 32 % of the words used in AFNN list are already included in SentiWordNet.

DATA COLLECTION & EXPERIMENTAL SETTING

We collected the 3698 Tweets regarding the San Bruno event in the first 24 hours from TwapperKeeper (www.twapperkeeper.com) which archived Tweets using the keywords #sanbrunofire and sanbrunofire. We used Crowdflower, www.crowdflower.com (Crowd Flower, 2011), a crowd sourcing system to manually classify these Tweets to generate a gold standard data set which was used to measure and benchmark our techniques.

SYSTEM ARCHITECTURE

Figure 1 shows the overall architecture of the sentiment detection system. We start by loading the sentiment lists AKNN, emoticons and the SentiWordNet. After loading the lists we match and search for the sentiment value of the words in Tweets as a result. Afterwards we express the sentiment in a mesh and classify them as positive, negative or neutral from the sentiment point of view. We divided the set of Tweets into two halves; where the Bayesian network was trained using half and tested using the other half. We used k-cross-validation to validate the statistical significance of our methods, for k=2; each experiment was repeated 10 times and the mean was presented. When there was conflict in judgment between the Bayesian and the other modules, we rank the result by the sentiment value calculated then we choose the class that has highest rank.

We start by preprocessing the Tweets. We remove the punctuation marks and the stop words. We use the list based on Lucene, the open source search engine. Then the tweets are tokenized. We used 50 % of the Tweets in order to train the Bayesian network for sentiment prediction. The next layer is responsible for matching the tokens of the tweets against the sentiment lists, emoticons, sentiment lists and SentiWordNet. The sentiment score for the Tweet is accumulated. The classification of the Bayesian classifier is added to the other scores that are accumulated by the three modules. Finally, a value for the sentiment of the Tweet is calculated depending on whether it is positive, negative or neutral.

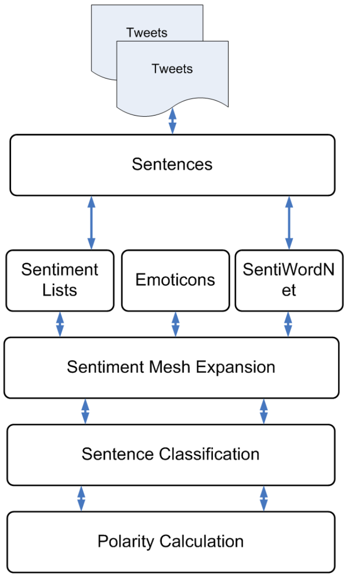

Figure 1. Reference Architecture and Overview

```
Input: Tweet
Output: SentimentValue
Initialize: Stop Word List, Emoticons, SentiWordNet, AKNN List
        Tweet=RemoveStopWords(Tweet);
        Tweet=StemTweet(Tweet);
        SentimentValue=0;
foreach Tweet
        foreach Tweet word
                SentimentValue= TweetSentiment[word]+SentimentValue
        end foreach
end foreach
BayesianResult=BaysianClassifier(Tweet)
GetSentimentValue(BayesianResult, SentimentValue)
```

Algorithm 1. Sentiment Detection Psuedocode

Limitations

Some of the Tweets were misclassified. Following are two examples of misclassified Tweets.

Tweet Number	Tweet Content	Value
1	Wow the entire top 10 trending topics of Twitter: San Francisco is all related to the #sanbrunofire pg&e natural gas line explosion	Neutral
2	I hate it when a tragedy happens and the news cameras show people that are smiling, waving and mugging for the camera. #SanBrunoFire #Fail	Positive

Table 3. Example of Misclassified Tweets

The first Tweet was considered neutral by our classifier. It is actually tricky to classify this Tweet even by a human being it can show astonishment due to the word wow but to classify it with either positive or negative sentiment is bit tricky. The second Tweet was classified positive. This can be attributed to having words like smiling, waving and mugging which when their sentiment value is added it turns out to be positive. To improve the accuracy of the classifier more training might help get it in better shape. Furthermore, performing more analysis on the structure of the sentence such as deducing the opinion holder and the entities that carry the sentiment might be helpful to improve the detection.

III. RESULTS

Comparison of Methods

We compare the sentiment detection techniques for disaster that we developed using precision, recall and the F-measure (Piwowarski, Gallinari, and Dupret, 1998). Precision can be defined as the fraction of retrieved documents that are relevant to the search. Recall is the fraction of the documents that are relevant to the query that are successfully retrieved. The F-measure combines precision and recall. The following equations describe how we calculated precision, recall and the combined F-measure (Egghe and Rousseau, 2007).

$$\Pr ecision = \frac{\mathrm{Re}levantTweets \cap \mathrm{Re}trievedTweets}{\mathrm{Re}trievedTweets}$$

$$\mathrm{Re}call = \frac{\mathrm{Re}levantTweets \cap \mathrm{Re}trievedTweets}{\mathrm{Re}levantTweets}$$

$$F - Measure = \frac{2.Percision.\mathrm{Re}call}{\Pr ecision + \mathrm{Re}call}$$

	AFNN + Emoticons + SentiWordNet	**Emoticons + SentiWordNet**	**Bayesian Network**	**AFNN+ Emoticons + SentiWordNet + Bayesian Network**
Recall	.95	0.93	0.72	0.96
Precision	.93	0.90	0.85	0.94
F -Measure	0.939	0.914	0.779	0.949

Table 3. Comparison of Techniques Used

Hypotheses Testing

Using the test method identified above we analyzed the San Bruno Fire Tweets to test two hypotheses about the sentiment of Tweets in disasters. Our first hypothesis stated the sentiment of Tweets changes over time in disaster recovery (time since the disaster). Table 3 presents results supporting this hypothesis. Table 3 shows the percentages of informative Tweets and sentiment measured values of information versus the sentiment value for both the 0-12 hour and 12-24 hour periods. The amount of information-based Tweets in the first 12 hours (61% of Tweets were informative) was more than the second half in which only 46% of the Tweets were informative. This suggests a need to share information during the early part of a disaster, replaced later by a need to express emotions. At the same time it is correlated with negative sentiment that shows the anger or the emotional negativity of participants.

Hours	**% Informative Tweets**	**% Sentiment Tweets**
0-12	61 %	39% (28 % with negative sentiment , 11 % with positive sentiment)
12-24	46 %	54% (18 % with negative sentiment , 36 % with positive sentiment)

Table 3. Measured Values of Information

On the other hand, the table shows the amount of sentiment based Tweets in the other half of the day changed to a more positive sentiment: the second half of the day includes 54% of emotional Tweets with higher positive sentiment than the first half of the day. This may be attributed to sympathizing with the victims dominating anger towards the organizations responsible for the disaster. In the first half day there is less information about such organizations or efforts to save victims. On the other hand as soon as there is enough information the sentiment is more positive. As a result, a good approach to address the lack of information is to set channels that can broadcast video content live for the places facing disasters whenever possible. The negative sentiment in the data set collected is 23 % while the positive sentiment is 23.5 %. The distribution of the positive and negative sentiment was not even; see (Table 3).

The second hypothesis we explored was: Information-based Tweets have more nouns than adjectives with low sentiment content. We found that the information-based Tweets have more nouns than adjectives with low sentiment content. Furthermore most of the information based Tweets that are classified as informative have a link to a website. We used the speech tagger to detect nouns, verbs and adjectives to calculate the information content in a post (Part of Speech Tagger, 2011). This aids in discovering the information based Tweets automatically based on the content of sentiment and that absence or existence of a link to a source of information. For the Tweets classified as informative Tweets, 81 % of them had links to an information source. The informative Tweets had 2.7 nouns compared to 1.2 adjectives on average. This can be considered as a way to increase the credibility of the post. Thus, we consider “information” to be identified as a part of speech (e.g., relative number of adjectives to nouns).

V. CONCLUSION & FURTHER RESEARCH

We presented a comparison of methods for evaluating sentiment in disaster microblogs and using the optimized system, explored patterns of change in emotion of the crowd during a disaster. The sentiment detection methods were derived from those used for documents, however, here we explored necessary modifications of the methods for the unique characteristics of Twitter microblogs, such as the shortness of the messages and use of emoticons etc. Because of these characteristics, not all the techniques can be ported as is. The main approaches that are used are list based and classification based. The limitation of these approaches is clear in their limited ability to expand the initial seed. Furthermore, there is a need to continuously maintain the list. We relied on an ensemble of techniques to characterize the sentiment of a Tweet. We used emoticons, sentimental ontology and frequent lists of sentiment words. We added further Bayesian networks to provide a way that can be flexible to categorize Tweets.

We then tested our method through analyzing the sentiment in 3698 disaster messages Tweeted in real time during the San Bruno fire disaster. The set was manually annotated to act as a gold metric using a crowdsourcing engine (Crowdflower). The method we developed proved to be effective at determining the sentiment of the Tweets. Using Bayesian Networks with SentiWordNet yielded the best precision and recall. Using Bayesian networks alone even with a training set equivalent to 50 % of the whole data yielded lower levels of accuracy for sentiment detection by 35 % compared with the combined approach.
In future work, we aim at extending our system by comparing it and including other classification techniques, and crafting the techniques as a service to be hosted in the cloud. This will enable for high load and throughput in near real-time. Furthermore, we would like to test and extend our system to other domains, such as politics and sports. Another enhancement that we plan to incorporate is integrating entity extractors to detect the sentiment towards a specific entity in a post which can give more details regarding the holders of sentiment and can give insights to Tweets that can be sarcastic which is a challenging task of Tweets. We would also like to explore incorporation of geographically tagged Tweets, which may impact the level of sentiment detection for people directly affected by the disaster.

ACKNOWLEDGMENTS

This work has been supported in part by the CMUSV Disaster Management Initiative and by a DMI affiliate, IntraPoint. We thank Dr. Martin Griss and Dr. Patricia Collins for advice and multiple reviews of the paper, and their many comments and changes that added tremendously to the quality of this paper.

REFERENCES

1. Agarwal, A., B. Xie, I. Vovsha, O. Rambow and R. Passonneau. Sentiment analysis of Twitter data. In Proceedings of the Workshop on Languages in Social Media, LSM '11, pages 30–38, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. 2011.

2. Baccianella, S., A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the Seventh conference on International Language Resources and Evaluation LREC10, pages 2200{2204. European Language Resources Association (ELRA), 2008.

3. Barbosa, L. and J. Feng. Robust sentiment detection on Twitter from biased and noisy data. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10, pages 36–44, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. 2010.

4. Bayesian Network Perl Implementation http://search.cpan.org/~kwilliams/Algorithm-NaiveBayes-0.04/lib/Algorithm/NaiveBayes.pm (Last accessed Dec 2011).

5. Bradley, M.M. and P.J. Lang. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida. 1999.

6. Crowd Flower http://crowdflower.com/ (Last accessed Dec 2011).

7. Egghe, L.and R. Rousseau. A theoretical study of recall and precision using a topological approach to information retrieval. Inf. Process. Manage., 34(2-3): 191-218.1998.

8. Hui, P. and M. Gregory. Quantifying sentiment and influence in blogspaces. In Proceedings of the First Workshop on Social Media Analytics, SOMA '10, pages 53–61, New York, NY, USA. ACM. 2010.

9. Jiang, L., M. Yu, M. Zhou, X. Liu, and T. Zhao. Target dependent Twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11, pages 151–160, Stroudsburg, PA, USA. Association for Computational Linguistics. 2011.

10. Miller G. A. WordNet: a lexical database for English. Commun. ACM 38, 11:39-41. 1995.

11. Nielsen, F. N. A new anew: Evaluation of a word list for sentiment analysis A in microblogs. Arxiv preprint arXiv:1103.2903. 2011.

12. Part of Speech Tagger http://search.cpan.org/dist/Lingua-EN-Tagger/Tagger.pm (Last accessed Dec 2011).

13. Pennacchiotti, M. and A. Popescu. Democrats, Republicans and Starbucks afficianados: user classification in Twitter. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11, pages 430–438, New York, NY, USA. ACM. 2011.

14. Piwowarski, B., P. Gallinari, and G. Dupret. Precision recall with user modeling (prum): Application to structured information retrieval. ACM Transactions on Information Systems (TOIS), 25(1):1. 2007.

15. SentiStrength http://sentistrength.wlv.ac.uk/ (Last accessed Dec 2011).

16. Tan, C., L. Lee, J. Tang, L. Jiang, M. Zhou and P.Li. User-level sentiment analysis incorporating social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11, pages 1397–1405, New York, NY. 2011.

17. Yu, C. T. and H. Mizuno. Two learning schemes in information retrieval. In Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '88), Yves Chiaramella (Ed.). pages 201-218. New York, NY, USA. ACM. 1998.