Sentiment Analysis of Twitter, SMS, and other Informal Short Text:

A Practical Survey of Challenges and Solutions

Melissa Serrano
Department of Computer Science and Engineering
Florida Atlantic University
Boca Raton, FL 33431
Mserra17@fau.edu

***Abstract*— Tweets[1], Text Messages (SMS), and other informally written short texts have become very popular in our society. Because of their brevity and informality mining these types of text are particularly challenging. However for these same reasons they provide a tremendous amount of information which can be useful to companies, such as the writers' opinion of a company's product, employee, or decision. Sentiment Analysis, sometimes referred to as opinion mining, refers to the process of computationally identifying and classifying a writer's opinion of a piece of text, in the case of this paper a 140-character or less text which is usually written informally - meaning the writer uses slang, misspelled words, and emoticons. Until recently, for the most part this type of sentiment analysis had not been researched, however over the past few years it has been explored in a fair amount of detail. In this paper I discuss sentiment analysis of various types of informally written short texts along with different approaches for feature reduction and pre-processing, sentiment classification, and analysis of entire messages or phrases within messages. In addition I briefly discuss a few successful sentiment analysis APIs (Application Programming Interfaces) that are being used for informally written short text. I will present the results of the techniques recently researched, as well as the techniques used by popular and successful APIs in order to provide the reader with a concise overview of successful techniques and expected results of particular approaches with regard to sentiment analysis of informally written short text.**

Keywords—Twitter, SMS, short text, sentiment analysis, opinion mining

I. Introduction

The goal of this paper is to provide the reader with a concise summary of techniques researched thus far and APIs which put them to use in the area of sentiment analysis of tweets, SMS, and other informally written short texts. The opinions in these types of text can be extremely useful for consumers, companies, and public figures. However it can be difficult to identify the polarity of the sentiment because of the informal context in which it is written.

In this paper I will discuss techniques used in analyzing the sentiment of informally written short text. The details of these techniques include feature reduction and pre-processing, and methods for sentiment classification which have been previously researched [1, 2, 3, 4, 5].

I will present a general framework for per-tweet sentiment analysis which first extracts tweets about a desired target subject [1], the use of emoticons as noisy labels in training data of Twitter messages [2], as well as discuss the exploration of sentiment analysis in an entire message or of a term within a message and how negation plays an important role in sentiment analysis [3].

I also present the results of tasks that other researchers have tackled. Since 2013, SemEval[2] has included sentiment analysis in Twitter[3] as one of the tasks. SemEval 2013 Task 2 challenged its' participants to determine the polarity of the sentiment within two subtasks: an expression-level task and a message-level task. In this paper I will provide a closer look at the results of these two subtasks which showed that the message-level sentiment analysis is actually more difficult, this is likely because an entire message can contain parts expressing both positive and negative sentiment [4]. SemEval 2014 Task 9 had participants who were tasked with the same challenge as in the related 2013 task but had the desire to test further by including Live Journal blog sentences along with the usual tweets and SMS messages, and also testing sarcastic tweets [5]. Currently underway, SemEval 2015 has expanded this task into three tasks related to sentiment analysis in Twitter; Task 1: Paraphrase and Semantic Similarity in Twitter, Task 10: Sentiment Analysis in Twitter, and Task 11: Sentiment Analysis of Figurative Language in Twitter [6]. Sentiment analysis in Twitter is also expected to be a task for SemEval 2016 but the specific tasks have not been released yet. SemEval provides an organized annual outlet aimed to inspire researchers of sentiment analysis to experiment with sentiment analysis systems in order to overcome the challenges in this area of research and therefore I have chosen to present the solutions proposed and results examined by the 2013 and 2014 SemEval participants of these relevant tasks.

I have reviewed some APIs which are in use for

[1]Tweets are 140-character limited text message sent through the online social networking website Twitter

[2]SemEval is an ongoing series of evaluations of computational semantic analysis systems http://en.wikipedia.org/wiki/SemEval

[3]Twitter is a social networking site http://en.wikipedia.org/wiki/Twitter

accomplishing the task of sentiment analysis of informally written short text. I will present these APIs while indicating which seem to be more popular and successful among developers at the present time as well as which use the techniques discussed in recently written papers to accomplish specific tasks [1, 2, 3, 4, 5].

II. Related Work

Over the past several years short informally written texts have become increasingly popular, whether they are tweets, SMS, or other microblogging platforms.

In 2008, just a couple of years after Twitter launched, Pang and Lee published a survey paper on the topic of opinion mining and sentiment analysis where they discuss techniques and approaches to directly enable opinion-oriented information-seeking systems [7]. Although the applications and datasets discussed are related to reviews, recommendations, business and government intelligence, and support of politicians or other public figures. These reviews and writings tend to be more formally written than the type of text I discuss in this paper. Nevertheless Pang and Lee's paper has been infamously referenced through research in the field of sentiment analysis. They take us through the main challenges also addressed (although in a relative manner to informally written short text) in papers written at a later date [1] such as: (1) extracting documents pertaining to a relevant topic, (2) extracting or identifying overall sentiment of the document, and (3) classifying and presenting the polarity of the sentiment identified in the document. They also address key concepts in classification and extraction which are also relevant in informally written short text such as classifying and extracting documents based on sentiment polarity while addressing negation, parts-of-speech, subjectivity, and topic relevance while building feature vectors or other representations.

More recently, there have been papers written which have analyzed sentiment analysis in Twitter as well as other informally written short text such as SMS and blog sentences [1, 2, 3]. [1] presented a cascaded classifier general framework for per-tweet processing (in contrast with batches of tweets which had been the usual form of processing the tweets). [2] Specifically used tweets with emoticons for their training set and then used a test set that did not necessarily contain emoticons.

The 2013 SemEval was the first year that SemEval tasked its participants with sentiment analysis in Twitter. This task addressed the issue of comparing approaches using suitable datasets and analyzed sentiment at both an expression-level and message-level [4]. One system used crowdsourcing on Amazon Mechanical Turk[4] to label a large Twitter training dataset along with additional test sets of both Twitter and SMS messages. The systems were tested with SMS to observe how effectively Twitter training data could apply to other informally written short text.

In the 2014 SemEval the task of Sentiment Analysis of Phrases within Short Texts took the 2013 SemEval task a step farther by testing the sentiment polarity on a message or phrase within a short piece of text, not only of tweets and SMS messages but also blog sentences [5]. The participants also targeted sarcastically written pieces of short text in their experiments comparing the consistency of the results with non-sarcastic messages and phrases.

III. Overview of Sentiment Analysis

Sentiment Analysis involves determining the polarity of an opinion expressed in a piece of text. The text can be interpreted as expressing positive, negative, or neutral sentiment. In the past sentiment analysis primarily targeted formally written text such as product or movie reviews. However since Twitter's launch in 2006 its number of active users has exploded, with over 500 million users in 2015. This, as well as other informally written short texts such as SMS, has provided another much desired outlet for extracting and analyzing users' sentiment. While many thoughts are exchanged through tweets and SMS, many users share their sentiment. However these types of text are usually not written as formally as reviews and therefore provide a greater challenge for detecting sentiment. Short text messages can often be a challenge even for a human to understand its sentiment.

The usefulness of sentiment analysis of informally written short text, especially those publicly available like tweets, is vast. Consumers can use the sentiment tweeted by others to decide if they want to buy a new product or watch a newly released movie. Companies can use it to monitor the feedback of their brand, business decision, or new employee. Political parties can use it to predict voter perception of candidates. The desire for extracting and retrieving tweets and other informally written short texts comes from the knowledge that these are casual messages and therefore will contain true, spur-of-the-moment thoughts and feelings which may be different if they were written in a well thought out review of the subject.

However because of the informality and brevity in which a tweet or SMS is written analyzing the sentiment of these types of texts provides greater challenges than sentiment analysis of more formally written pieces of text.

Through the rest of this paper we will examine the challenges in classification and extraction posed by these types of informally written short text, compare the experiment results of suggested solutions of these challenges, and then look at APIs that are currently implementing the successful solutions discussed herein.

IV. Challenges

When writing a tweet or SMS most people don't use proper grammar or punctuation. In fact, they often misspell words, use slang, or even express their thoughts or emotions with emoticons. This is the primary challenge when attempting to analyze the sentiment of any informally written short text. Another challenge is the limited character length of these types of texts. Tweets are limited to 140-character length and most SMS are limited to 160-character length. Previous experiment results [4] have shown that the fact that the length of a piece of text is short may not actually be the challenge. The challenge was first for the writer, to be able to fit the words they want to express within the character limits. Some writers might not be

[4]Amazon Mechanical Turk https://www.mturk.com/mturk/welcome

as crafty as others resulting in a message that loses the intended meaning, and the crafty writer may make it challenging to computationally read the message because of acronyms or shortened words resulting in awkward misspellings. It can also be difficult to detect sarcasm in these types of short texts. With the text being only one or two sentences when sarcasm is used it is difficult to determine under which polarity the sentiment should be classified. As we will see by solutions proposed and experiments executed techniques can be applied to feature reduction and pre-processing, as well as sentiment classification methods to address these challenges.

V. Solutions

Let's examine solutions to the challenges we face when attempting to analyze the sentiment of informally written short text by breaking the process up into tasks: features and pre-processing, as well as sentiment classification methods. I will present solutions proposed and experiments executed within each of these tasks to address these challenges.

As a research community we want to be able to analyze the sentiment of any informally written short text. As pointed out in [2], by using the Twitter API we can access millions of tweets to use for training. So the trend among experiments is to train data with only tweets and test data with tweets, SMS, and other microblogging sources.

Since tweets are used for training sets some common pre-processing is done. The Twitter language model has many unique properties that can be used strategically.

Since the hashtag symbol prepended to a word is used to categorize a tweet into a particular subject, it has been observed and used [3] in sentiment analysis systems as signals of positive and negative sentiment. (e.g., *#loveIt*, *#hateIt*, indicates positive and negative sentiment respectively)

Because of the character limitation per tweet many users express their sentiment with emoticons. The Twitter API allows extraction of tweets of a particular sentiment, mapping a number of emoticons to either positive or negative as used in [2] and displayed in Table 1.

Table 1: List of Emoticons [2]

Emoticons mapped to :)	Emoticons mapped to :(
:)	:(
:-)	:-(
:)	: (
:D	
=)	

A. *Features and Pre-Processing*

As [2] reduced their features Go et al. used the @ to indicate Twitter usernames and replaced all of these with an equivalence class token (USERNAME). Likewise they did the same for all links found in tweets, replacing them with an equivalence class token (URL).

A clever observation made in [2] is that sometimes misspellings in tweets are not necessarily to shorten the tweet but to show exaggeration. (e.g. I looooooooove McDonalds!) The issue is that if you search for "love" this tweet will not be returned. To address this type of misspelling Go et al. pre-processed words so that any letter occurring more than two times in a row would be replaced with exactly two occurrences of the letter.

For SemEval-2014 participants created a system [5] whose main focus is to evaluate sentiment polarity of messages as well as phrases within messages, and also evaluate the test set on sarcastic tweets. While they used features common to other systems which we previously discussed they found that prior polarities of words to be one of the most important features in sentiment analysis of phrases. Another feature they implemented which was not necessarily commonly implemented among other experimental systems, their system clustered multiple misspellings of a word together (e.g., *anyone, anybody, any1, ne1).*

B. *Sentiment Classification*

Common machine learning algorithms used in sentiment classification were Naïve Bayes, Maximum Entropy, and Support Vector Machines (SVM).

[2] Showed that these commonly used machine learning algorithms have above 80% accuracy when trained with emoticon data.

The system discussed in [5] which participated in SemEval 2014 experimented with these various machine learning classifier, but decided to use Support Vector Machines (SVM) in their final system model because they found it outperformed all other classifiers.

The experiments executed in [1] also made the observation of SVM outperforming other machine learning classifiers, which was consistent with that of SemEval 2014 participants [5]. Another notable observation made in [1] is that a Multinomial Naïve Bayes (MNB) outperformed other variants of Naïve Bayes. Also when building their input vectors they experimented with both word-count and binary implementations. However, most tweets or SMS messages are primarily binary, because of their character limitations most words won't be used more than once.

C. *NotesCommon Among Experiments*

1) Training sets were created with test sets targeting tweets written only in English. This should be noted for future research because Twitter is a multi-lingual application, as of the date of this writing Twitter supports 33 languages. Training and testing data in other languages will certainly pose unique challenges which may not be addressed through English language experiments.

2) Part-Of-Speech (POS) is not a helpful feature for informal short text classification. Consistent with previous works [2] observed that POS tags were not useful. Go et al. noted that the use of POS tags decreased accuracy when using Naïve Bayes and SVM.

VI. APIs

The most practical and useful API related to sentiment analysis in Twitter is Sentiment 140 accessible at http://www.sentiment140.com/. It was created by Stanford students and authors of [2] that provided us with great insight into useful features to use and appropriate pre-processing. So if a reader plans to analyze sentiment on Twitter this API is the key tool to use, also accompanied by the reading of [2] as they discuss in greater detail than in this paper their experiments and findings which they applied in the creation of this API [8]. They note in the General Information page of the API that because of the observations made during the experiments and writing of [2] they added new pieces to improve their system which are not described in the paper.

There have been many other APIs developed recently to target this growing area of sentiment analysis. A common one among developers seems to be Semantria, available at https://semantria.com/. The API offers 15 languages and was put to the test by another group of Stanford Computer Science students who used Semantria to test the validity of an established model on the process of opinion formation [9].

VII. Conclusion

In summary, I have presented you with an overview of previous research, experiments, and results related to sentiment analysis in tweets, SMS, and other informally written short text. I discussed the challenges faced when attempting to analyze the sentiment of these types of text, as well as present some solutions used in experiments. I then point out two successful APIs that are in use by developers and researchers to analyze the sentiment of short text, specifically of Tweets and FaceBook posts [9]. Where one API, Sentmient 140, directly relates to a paper we discussed [2]. This paper should serve as a practical overview of the challenges, solutions, and accessible way for developers and researchers to understand and apply sentiment analysis to informally written short text.

References

[1] Ameeta Asiaee T., Mariano Tepper, Arindam Banerjee, and Guillermo Sapiro. 2012. If you are happy and you know it… tweet. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 1602-1606, New York, NY, USA. ACM.

[2] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N ProjectReport, Stanford*, pages 1-6.

[3] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M Mohammad. 2014. Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research*, 50:(723-762).

[4] Theresa Wilson, Zornitsa Kozareva, Preslav Nakov, Sara Rosenthal, Veselin Stoyanov, and Alan Ritter. 2013. Semeval-2013 task 2: Sentiment analysis in twitter. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval-2013)*, June: (312-320).

[5] Ameeta Agrawal and Aijun An. 2014. Kea: Sentiment Analysis of Phrases Within Short Texts. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval-2014)*, August: (380-384).

[6] SemEval-2015: Semantic Evaluation Exercises. *International Workshop on Semantic Evaluation (SemEval-2015).* http://alt.qcri.org/semeval2015/index.php?id=tasks

[7] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1-135, 2008.

[8] Google. Sentiment140. (2012). http://help.sentiment140.com/home

[9] Semantria LLC. Semantria Out-of-the-Box Reliability. (2015). https://semantria.com/case-studies