INTERNATIONAL JOURNAL OF SOCIAL Sciences and Information Technology
ISSN 2412-0294 www.ijssit.com

http://www.ijssit.com Vol II Issue IV, June 2016
ISSN 2412-0294

LIMITATIONS OF SENTIMENT ANALYSIS ON FACEBOOK DATA

[1] **Sidney B. Omondi Ochieng**
Data Scientist,
iHub Research Lab
sidney@ihub.co.ke

[2] **Michael Loki**
Research fellow,
iHub Research Lab
mkloki@gmail.com

[3] **Nanjira Sambuli**
Data Scientist,
iHub Research Lab
nanjira@ihub.co.ke

Abstract

Sentiment analysis of short texts such as single sentences and Twitter messages is challenging because of the limited contextual information that they normally contain. Effectively solving this task requires strategies that combine the small text content with prior knowledge and use more than just bag-of-words. This survey covers experiments done as part of ongoing Umati Project at iHub research lab in Nairobi, Kenya and published literature. This study focuses on challenges of analyzing data on Facebook as compared to Twitter. Three established limitations; heterogeneous nature of news group; time series analysis and association of words, have been discussed.

Keywords: *Bag-of-Words, Natural Language Processing, Sentiment Analysis, Umati Project*

Introduction

With the advent of user generated content, usability and interoperability of web platforms, people have become more eager to express and share their opinions on web regarding day-to-day activities and global issues as well. Evolution of social media has also contributed immensely to these activities, thereby providing a transparent platform to share views across the world [2]. As part of the ongoing Umati Project from iHub research lab in Nairobi, we took time to examine how Twitter users responded to the unfortunate attack on Garissa University in 2015. Garissa University is a public university in the Nothern part of Kenya which suffred an attack by a known terrorist group from a neighboring country.

Introducing Umatex

We used a modified bag-of-words model to analyse the data and create a tool called *"Umatex"*, which acts as a filter for dangerous speech. The bag-of-words model is a simplified representation used in Natural Language Processing; a text is represented as a bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity. Multiplicity is, in simpler terms, the number of times a word appears in a set of documents. *Umatex*, therefore, is a module within the Intelligent Umati Monitor that removes noise from collected hate speech statements and ranks them accordingly.

Umatex is proving to be a valuable tool in reducing the workload of data coders. However, additional work to improve it is needed, and ongoing. We are also working on expanding the bags-of-words and creating new ones. There also are other techniques that can be used to increase accuracy. All this needs to be benchmarked against human coders and be tested on multiple datasets, an ongoing process.

Background of study

Sentiment analysis aims to study the opinions and emotional aspects of individuals on given texts. Measuring influence on social media is a big business and relatively a new dimension of analyzing sentiments. Smart companies have taken advantage of social media to identify opinions for a diverse disperses fields such as business intelligence, healthcare, customer journeys, sales generation, sentiment analysis, e-learning, political science, web analytics etc… all in real-time for quick and calculated decision making.

With a large number of people embarking on a trend of actively voicing out their opinion online on social networks and forums, social media, e.g., Twitter, Facebook, have become a major source for social data mining.

Empirical Evaluation

Sentiment Analysis Limitations and Techniques to Improve Results

While computer systems and machine learning algorithms are getting better all the time, they still face challenges when deciphering human sentiments in online statements.

Examples where sentiment analysis tools fall short:

- Irony, humor and other subtleties of human speech, like how the emoticon 🙂 can change the tone of an otherwise negative statement.
- Spam-loaded conversations in social media that strike people as inauthentic.
- False negatives, where the software sees a negative word like “crap” but doesn’t realize it’s positive in the overall context—”Holy crap! I loved this!”
- Cultural differences, where some people from some countries might be more or less effusive in their use of language.

Techniques that help improve the effectiveness of sentiment analysis:

- Picking a limited number of concrete product features to analyze
- Pairing sentiment analysis tools with human analysts to examine contextual references
- Use sentiment analysis as a starting point to identify issues for follow-up action
- Connect sentiment analysis questions to a business problem
- Going beyond the polarity of “positive” and “negative” to classify sentiment, and using more fine-grained categories like “angry,” “happy,” “frustrated,” and “sad.”

Baseline Approach

To create *Umatex*, hate speech statements collected from the first phase of the Umati project were analyzed, based on the hypothesis that there were certain features common to the all the statements collected. These common features were then examined vis-a-vis the Umati framework for categorizing online inflammatory speech. According to the Umati framework, a dangerous speech statement:

1. targets at a group of people based on their common affiliation and not a single person

ISSN 2412-0294

2. may contain one of the hallmarks/pillars of dangerous speech
3. contains a call to action

Targets a group of people and not a single person

Dangerous speech is harmful speech that calls on the audience to condone or take part in violent acts against a group of people. Such speech is directed at a group, or at a person as part of a group: a tribe, religion, etc.

It is important to note that an ugly or critical comment about an individual - a politician, for example - is not hate or dangerous speech unless it targets that person as a member of a group. As noted in our previous reports, during emotive periods, it is not uncommon for negative statements to be made against politicians and other influential personalities.

With this in mind, a bag-of-words was created for certain categories under which people are grouped in Kenya: tribe, political affiliation, religion, region of origin and sexual orientation. Therefore, the 'tribe' bag for example looked for words (and their variations) like *Kamba*, *Kikuyu*, *Luhya*, *Luo*, *Kisii*, *Kalenjin*, *Giriama*, *Somali*, etc. while the 'political affiliation' bag has the words (and their variations) like '*Jubilee*', '*ODM*', '*PNU*', '*CORDian*', '*CORDed*', '*Chupilee*' etc. Each bag has a particular weighting. This means that any tweet or Facebook post which contained any word in the tribe bag as well as one or more other bags could be considered inflammatory provided their combined weight passes a certain threshold.

May contain one of the hallmarks/pillars of dangerous speech

Three hallmarks common in dangerous speech statements are:

- Comparing a group of people with animals, insects or vermin
- Suggesting that the audience faces a serious threat or violence from another group ("accusation in a mirror")
- Suggesting that some people from another group are spoiling the purity or integrity of the speakers' group.

Of these three, it is was easiest to build a bag-of-words for the first hallmark - comparing a group with animals, insects or vermin. Given the highly contextual nature of the other two hallmarks, it would be difficult to use the same model. It is, however, not impossible and will be an avenue explored in future to make the algorithm better.

Contains a call to action

Dangerous speech often encourages the audience to condone or commit violent acts on the targeted group. The six calls to action common in dangerous speech are calls to:

- discriminate
- loot
- riot
- beat
- forcefully evict
- kill

While building a bag-of-words for this category would seem straightforward, there are various nuances of language and context to consider. There are numerous ways to show discrimination or to make a call to loot, to beat or kill in the various languages used in Kenya. For example, taking the word *kill*, you could have *destroy*, *rid*, *massacre*, *execute*, *terminate* and, one that is sometimes used in Kenya, *finish*.

Each bag-of-words category created is assigned a weight using data from the first phase. If a word from a bag appears in a sentence, the weight of that bag is added to the overall weight of the sentence; for example, if the bag has a weight of 0.5 and a sentence contains 3 words from that bag the weight of that sentence will be 0.5*3 = 1.5. This is done for each bag and the total weight of the sentence will be the sum of the weights from all bags.

If the sentence meets a certain threshold weight, which is currently at 4, it is then considered to be potentially dangerous speech, everything else is dropped as noise.

The method described above is used to filter out noise. It is not a method that will automatically lead to identifying dangerous speech texts. Rather, it is a method to be used along with human input; a human would still have to manually go through the text omitted by *Umatex*, for instance, to ensure that significant data is not filtered out altogether. Speech, in general, is highly contextual, and context is something that is difficult to teach a computer. As we have found and previously noted, dangerous speech is both an art and a science.

The purpose of *Umatex* is to help reduce the workload of human data coders. In Umati Phase II, we have collected several gigabytes of data; millions of individual pieces of text and related metadata. It would be fairly expensive, not to mention inefficient, to hire annotators to go through it all. *Umatex* is able to quickly and efficiently sort through this data reducing its size by a factor of more than 10, while guaranteeing that a certain percentage of dangerous speech in this use case remains in the filtered text (This percentage is currently 70% based on tests against data collected in Umati phase 1) . Human coders will then be able to go through this reduced dataset to monitor for dangerous speech.

From Facebook data collected around the Garissa attack, there were statements of discrimination against people of the Islamic faith and what may be considered a call to evict them by destroying their places of worship. There was a direct call to evict and a comparison of a group of people to animals. Some of the statements took a tribal tilt, more specifically, “accusations in a mirror” that is, the suggestion that one group faces a threat from another.

Findings

In previous works, we looked at the sentiment of tweets around the attack, some users were driving conversation, the nature/content of the conversations, and made some inferences about the audiences engaging on the topic. However, when looking at data off Facebook, which was collected from pages and groups, we had to apply different methods and approaches to analyze the same data sets. Some of the limitations that we encountered while analyzing data off Facebook are;

1. ***Heterogeneous nature of News groups***

Communication on Twitter differs from Facebook in various ways. Key among them is how conversations occur around a particular topic. Thus, a new tweet can be in reaction to any previous tweet on the timeline and the conversation around a hashtag is largely homogeneous and continuous. On Facebook, however, conversations manifest around posts in the form of comments and replies. Individual posts are independent of one another, thus comments and replies around one post are rarely in reaction to a different post (or if they are, this can only be derived from having the context, or if a new post is tagged to a previous one. These are not accessible on one continuous timeline of events as in the case of Twitter).

2. *Time series analysis*

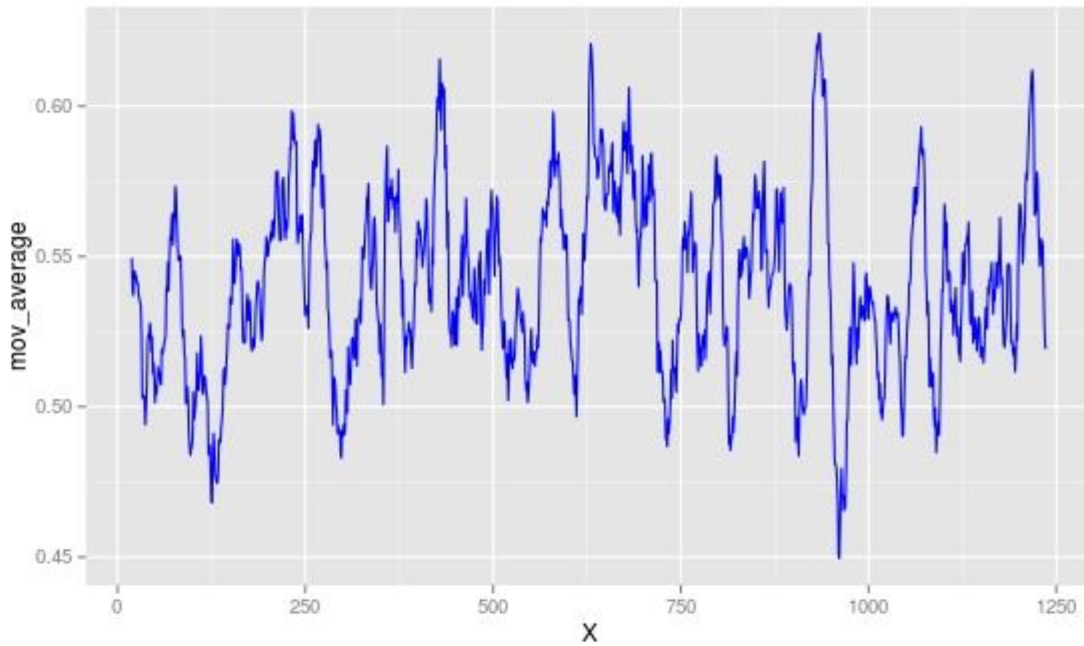

Figure 1.0 above shows the moving average for Facebook as compared to twitter graph 2.0 below

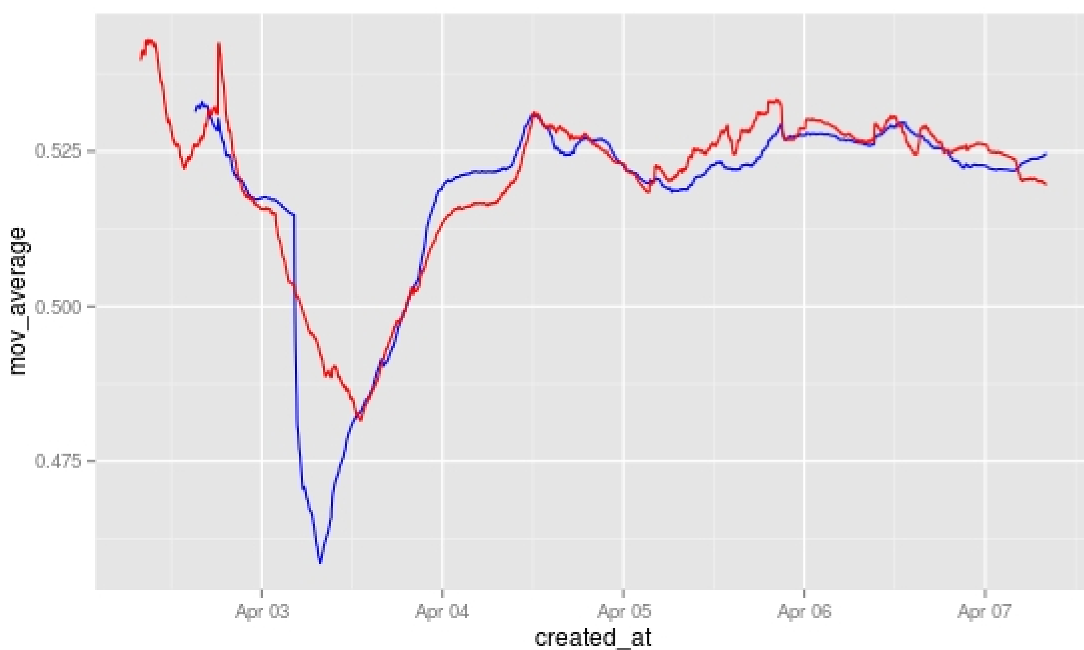

Fig 2.0

The many and sudden changes in the sentiment indicate that doing a time series analysis on Facebook data to find dangerous speech is difficult, complex and subject to a lot of noise while the Twitter graph has a distinct dip that make it easy to analyze.

3. Association of words

Umatex was also run on Twitter and in line with the conversational analysis from part 1 some of the dangerous speech found was from accounts not associated with Kenyans. Unlike Facebook conversation from this dataset, on Twitter our data shows that it mostly focuses on discrimination of the Muslim and Somali community. Also unlike Facebook all the dangerous speech was in English.

Most of the tweets got no reactions in the form of retweets(amplification) or replies, other than one which got an outsized reaction of 800+ retweets and several replies including what we define as KOT(Kenyans on Twitter) cuffing.

Related Work

Most work in sentiment analysis has focused on identifying positive or negative sentiment in text passages online. These studies can be broadly classified into two categories: knowledge-based approaches and learning-based approaches.

Knowledge-based approaches primarily use linguistic models or other forms of background knowledge to classify the sentiment of passages. A large focus of this area is the use and generation of dictionaries capturing the sentiment of words. These methods range from manual approaches of developing domain-dependent lexicons [7] to semi-automated approaches [5, 4, 8, 3], and even an almost fully automated approach [6]. As observed by Ng et al. [9], most semi-automated approaches yield unsatisfactory lexicons, with either high coverage and low precision or vice versa. More recently, Pang et al. [1] successfully applied a machine learning approach to classifying sentiment for movie reviews. They cast the problem as a text classification task, using a bag-of-words representation of each movie review. They demonstrate that a learning approach performs better than simply counting the positive and negative sentiment terms using a hand-crafted dictionary.

Labeling blog posts as positive or negative is very complex, even for humans. While efforts are made to focus solely on the post content when categorizing posts, it is still the case that comments, citations, and quotes from other sources are included in the main body of a post. When a blog post's page becomes an area of discussion on a certain subject, labeling the entire page as positive or negative can be quite difficult. This is especially true for political blogs, where writers often make comparisons between multiple candidates, policies, or events. The issue of sentiment analysis is further complicated by the fact that bloggers often use jokes, anecdotes, and cultural references to illustrate their opinions, making the labeling task unclear for people unfamiliar with the relevant facts or references. This makes sentiment classification extremely difficult for most algorithms [10].

References

[1] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In EMNLP, 2002.

[2] http://data-informed.com/improving-effectiveness-of-customer-sentiment-analysis - accessed on 22 February 2016

[3] H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In EMNLP, 2003.

[4] L. Zhuang, F. Jing, and X.-Y. Zhu. Movie review mining and summarization. In CIKM, 2006

[5] M. Hu and B. Liu. Mining and summarizing customer reviews. In KDD, 2004.

[6] P. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. ACL, 2002.

[7] S. Das and M. Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Asia Pacific Finance Association, 2001.

[8] S.-M. Kim and E. Hovy. Determining the sentiment of opinions. In COLING, 2004.

[9] V. Ng, S. Dasgupta, and S. M. N. Arifin. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In ACL, 2006.

[10] P. Melville, W. Gryc. And R. Lawrence. Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification, 2014.