Mining voter sentiments from Twitter data for the 2016 Uganda Presidential elections

IJOTM
ISSN 2518-8623

Volume 3. Issue II
p. 12, Dec 2018
http://jotm.utamu.ac.ug
email: ijotm@utamu.ac.ug

Isaac Mukonyezi
imukonyezi@utamu.ac.ug
Uganda Technology and Management University

Claire Babirye
Uganda Technology and Management University

Ernest Mwebaze
Uganda Technology and Management University

Abstract

Micro-bogging platforms like Twitter have proved to be fertile ground for political campaigns. In several applications, Twitter data has been mined to understand sentiments of the population, to determine trends and to understand the informal communication amongst people. A key challenge with the Twitter platform is that it produces immense quantities of noisy unstructured user-generated data across multiple social relations. The uniqueness of social media data calls for novel data mining techniques that can effectively handle user-generated content with rich social relations in order to build descriptive and predictive models of social interactions. We analyse Twitter data for the #UgandaDecides hash tag for Uganda Presidential elections 2016, during the period of January to February. We derive inferences from the data and show that in some cases Twitter can be informative on actual events happening on ground. In the analysis we use a word-emotion lexicon to determine the nature of sentiments in the tweets and semantic orientation, to determine the ties conversations within the tweets have with positive and negative contexts, this is based on the pointwise mutual information technique. We find that twitter data analytics using both intensity-based measures and sentiment analysis can be useful to reflect the current offline political sentiment and we make a number of observations related to the task of monitoring public sentiments during an election campaign, including examining a variety of sample sizes, time parameters as well as methods for quantitatively and qualitatively exploring the underlying content.

Key words: *Text mining, Twitter Analytics, Information Retrieval*

Introduction

Big data, an evolving term used to describe the exponential growth and availability of data, both structured and unstructured that has the potential to be mined for information; has played a significant impact in informing decision making as well as contributing to making information transparent and usable.

Big data can be used to predict the future through the use of predictive modelling to come up with possible results. For example for the US 2008 presidential elections the key role played by big data in these elections is no secret, as it jumped into the spotlight when the Senator Barack Obama invested significant resources into analysing big data, which helped propel him into the white house. At the same time, Nate Silver; an American statistician used big data to project Obama's presidential victory with a stupefying accuracy, correctly predicting the outcomes in 49 of the 50 states in the 2008 presidential election.

Social media has become a multifunctional platform, giving users the opportunity to express, share and influence. Social media platforms like the micro blogging service, Twitter has evidenced a significant influence on information diffusion such as Mr. Museveni not attending the first Uganda Presidential debate ever, spread in the world.

On the 18th February, 2016, the republic of Uganda decided on a new elect president for a 5 year term through the national voting that was carried out then. However, before the Election Day (from November 2015), social media sites (for instance Facebook, YouTube and twitter) were being used by the different presidential candidates for political campaigns and also to engage with the citizens. The activities on social media about the political atmosphere became hyperactive as the nation was tending towards elections. For instance, on twitter, this was evidenced by the different hash tags that were created; #UgandaDecides explicitly showing that political issues were clearly on the minds of many users. Furthermore, politicians were trying to mobilize supporters on their user profiles for example the two top candidates @KagutaMuseveni and @KizzaBesigye1. While some political analysts are already turning to the 'Twitter sphere' as an indicator of political opinions, others have suggested that the majority of the messages are a 'pointless babble'. As a result, we aim at answering the question whether the microblogging platform; twitter can actually help in prediction of election results.

For this study we analyse 80,000 tweets regarding the presidential elections posted under the #UgandaDecides hash tag. To be formal we analysed a sample of tweets in a given time interval [t1 and t2] where t1 (i.e. '17.02.2016 00:00 UTC') and t2 (i.e. '18/02/2016 00:00 UTC') denote the time stamps for the first and last tweets respectively, collected by querying the Twitter API. Analysis is also done on the election results data that were released by the Uganda Electoral Commission to determine existence of any correlation between the results and the sentiments from the twitter data.

We visualize these tweets in different ways as sentiments to find out people's opinions; as timelines where we draw tweets on a bar chart to depict the number of tweets posted at different times on the pre-election days; as expression of themes, for which we draw a word cloud to show the top hashtags mentioned in the different tweets during the election period; and as a communication network, where we discover the top tweets and the top accounts that contributed to the conversations in the dataset. For proper analysis and understanding of the election results, we also visualize the results on maps to show the distribution of votes across the nation for the best presidential candidate.

The rest of this paper is organised as follows: in section 3 we discuss recent work that has been done in this field, then in section 4 we discuss the different steps we go through in collecting the data, pre-processing it and then the different techniques we apply in analysing the data. We also conclude by looking at the sentiments in the twitter data.

Background

Twitter is very suitable for news information, publishing and propagation. The text of tweet is short, so it is easy for users to post their tweets with mobile phones or other tools and as a result, users can publish what they see in real-time, especially for the news information.

Twitter is one such service that allows users to broadcast short textual messages of up to 140 characters to an audience of followers using web or mobile based platforms (Tumasjan, 2010). The real-time nature of twitter accommodates users' ability to post events as they happen and thus positively affects information diffusion in the media as individuals share information and post their reactions on the events in topic as they happen. Such information is highly valuable, since it offers a user perspective of events that would otherwise not be available. As a result of its growing ubiquity, speed and cross-platform accessibility, twitter is increasingly being considered as a means for effective communication for the most trending national events.

Twitter employs a social model called 'following' in which the user is allowed to choose any other users he wants to follow without any permission or reciprocating by following her back (Asur, 2010; Tumasjan, 2010). The one she follows is her friend and she is the follower. Being a follower on twitter means that she receives all the updates of her friends. The 'following function' makes it very suitable for news information diffusion. When a user participates in a new topic discussion, the related tweets will be pushed to the followers, if the followers are also interested in the topic, they may take part in the discussion too. Thus, the new topic diffuses very fast (Jahanbakhsh, 2014).

Tweets appear in three forms: normal tweets, replies, and retweets. Normal tweets are the tweets except replies and retweets. Replies are messages to any user with a sign '@' followed by the target user's name at the beginning of the tweet. Retweets are messages which users share from friends to their followers which begin with the sign 'RT'. Users can also add tags for the tweets that is to say 'hashtags' as known on twitter. They begin with a '#' and they are created basing on topic of focus. The different forms of tweets and hashtags make twitter data stream more readable and understandable. Tweets appear in form of short text, so that in analysing the tweets we do text data mining to derive high quality information from the tweets which appear unstructured. For proper analysis of the texts we use one of the most recent techniques referred to as sentiment analysis to derive the opinion or attitude of the writer whether positive, negative or neutral (Tumasjan, 2010).

Related Work

Social Media

Social media is taken as a generic term for social interactions built on a multitude of digital media and technologies, which allow users to create and share content. Major social media sites are popular venues for publishing rich and diverse information about a variety of real-world events (Wang, 2012). An example of such event is the 2016 series of presidential debates in Uganda. As a result of their flexibility and ease of use, social media have recently emerged as a popular medium for providing new sources of information and rapid communications.

However the valuable user-contributed information is not readily available in structured form, since the social media content is noisy and highly heterogeneous (Asur, 2010). Information on social media sites appears in unstructured natural language text form; it contains abbreviations and truncated phrases, thus to extract meaningful indices from the text and thus make the information contained in the text accessible to the various data mining algorithms that is to say statistical and machine learning algorithms, text mining is carried out.

We do social media mining, a process that involves extraction, analysis and representation of informative patterns from social media data to extract highly valuable information from the text data. Information extraction is the task of finding structured information from unstructured or semi-structured text. It is an important task in text mining and has been extensively studied in various research communities including natural language processing, information retrieval and web mining. Two fundamental tasks of information extraction are named entity recognition (Ritter, 2011) and relation extraction. Entity recognition refers to finding names of entities such as people, organizations and locations and relation extraction refers to finding the semantic relations between entities.

Opinion Mining

An opinion is simply a positive or negative sentiment, attitude, emotion or appraisal about an entity or an aspect of the entity from an opinion holder (the author of the posting which carries the opinion) (Liu, 2012. The attitude or sentiment of an opinion, can be either positive, negative or neutral. These are generalized as sentiment orientations (also referred to as polarities, opinion orientations, and semantic in orientations).

An example of an opinion is; Claire bought a Tecno Phantom 5 last year in May. It was such a nice phone with a finger print scanner at the back just below the 13MP camera. The phone has an inbuilt storage space of up to 32GB. Much as the battery life was not so long, I found it worth the purchase. However, my colleagues were jealous when I acquired the phone, and they kept on saying 'it's too costly.'

Liu (2012) suggests that an opinion be construed as a quintuple (a, b, c, d, e) where a is the name of an entity, b is an aspect of the entity a, c is the orientation of the opinion about the aspect b of an entity a, d is the opinion holder and e is the time when the opinion is expressed. The opinion orientation c, can be positive, negative or neutral, or be expressed with different strength/intensity levels.

From such an example, a is 'Tecno Phantom 5 (the name of the entity), b is the battery life, c is positive (It was such a nice phone with a finger print scanner at the back just below the 13MP. The phone has inbuilt storage space of up to 32GB), negative (battery life not so long, it's too costly), d is Claire and the friends; e is May last year.

Opinions are central to almost all human activities because they are key influencers of our behaviours (Lei, 2012). Peoples' thoughts, opinions and ideas have always been an important piece of information for most of us during the decision-making process. Indeed according to a survey by (Ritterman, 2009), Twitter data provides more than factual information about public opinions on a specific topic, yielding better results than information from the prediction market (Asur, 2010). The internet and the web have made it possible to find out about public opinions and experiences through the social media sites (for instance twitter) which provide a platform for public sharing of information.

(Liu, 2012) Articulates that the opinionated postings in social media helped reshape business and sway public sentiments and emotions, which profoundly impacted on the social and political systems. Further, the postings also mobilized masses for political changes such as those that happened in some Arab countries in 2011 (Buche, 2013).

Opinions need to be studied from a large number of opinion holders, one opinion from a single person is usually not sufficient for action (Liu, 2012); thus more opinions from a sample are desirable so as to derive inferences on a population.

Data Analytics

In this section we discuss the number of tweets, how widely the information diffused, who the top content contributors were and how the general social network public perceived the entire electoral process.

Data Mining

The tweets collected are structured and are all uniform with the following key attributes:

- `text`: the text of the tweet
- `created`: the date of creation
- `favoriteCount`, retweetCount: the number of favourites and retweets
- `favorited`, retweeted: whether the user has marked the tweet as favourite or retweeted the tweet
- `id`: the tweet identifier
- `Latitude and Longitude`, geo-location information if available
- `screenName`: the twitter name of the user
- `replyTo`: user identifier if the tweet is a reply to a specific user
- `isRetweet`: identifier if the tweet is a retweet or not

A sample tweet will follow a Json structure as follows {"**created_at**":"Sun Jan 15 21:10:49 +0000 2017","**id**":820740040842219525, "**text**":"@Uganda This is the best weather to enjoy a #holiday", "**source**":,"**location**":null,"**time_zone**":"Africa/Kampala","**geo_enabled**":true,"**lang**":"en","**geo**":null,"**coordinates**":null,"**place**":null,"**retweet_count**":0,"**favorite_count**":0,"**entities**":{"**hashtags**":[holiday],"urls":[],"user_mentions":[{"screen_name":"Uganda","favorited":false,"retweeted":false, "**screen_name**":"imukonyezi","**name**": "Mukonyezi Isaac","**id**":78406161}.

With this data we can therefore check who is most favourited/retweeted, who is discussing most, and who's discussing with whom, what are the most popular hashtags and most importantly we can look at the text of the tweet to find out patterns, sentiments, word orientations and co-occurrence of words.
We start our data mining by performing pre-processing techniques like breaking the text down into words through a process known as tokenisation which is one of the most important steps in text analysis. The purpose of tokenisation is to split a stream of text into smaller units called tokens, usually words or phrases (Russell, 2013).

We also improve the richness of the pre-processing by removing regular expressions from the texts. In every language, some words are particularly common. While their use in the language is crucial, they don't usually convey a particular meaning, especially if taken out of context. This is the case of articles, conjunctions, some adverbs, etc. which are commonly called stop-words. We use the existent library of stop words in Natural Language Tool Kit (NLTK) for the English language and also build a custom list of stop-words. After pre-processing the data we go ahead and apply some text analysis.

The number of tweets collected during Election Day and the day before, including retweets is shown in Table 1, indicating a steady rise in the use of Twitter to share and discuss on the elections.

Table 1: Total Collected Tweets

Election Day	Pre-Election Day
61,906	23,191

In Table 2 we tally audience reactions to the original tweets (excluding retweets), showing that most of the tweets on both days didn't spark a lot of reaction, but notably 26% of all natural or original tweets collected were retweeted and favorited at least once for both the pre-election and election days.

Table 2: Audience Reaction to Tweets

Election Day (14,283 Original)	**Not Retweeted**	**Retweeted**
Not Favorited	6848 (48%)	2210 (15%)
Favorited	1551 (11%)	3674 (26%)
Pre-Election Day (6,005 Original)		
Not Favorited	3081 (51%)	847 (14%)
Favorited	525 (9%)	1552 (26%)

We also represent the time distribution of the tweets, showing that the number of tweets shared across the day increased significantly on the Election Day as compared to the previous day. This provides us with information to know at what times of the day people were active on social media and what are the topics driving those conversations and derive some inferences on the data. There were more tweets posted on the election day than the two days before the elections, although due to the directive to turn off social media on the day of election indicated by a poor number of tweets in the earlier morning and rising through the day and maintaining the general trend from the previous days due to the use of Virtual Private Network (VPN) applications.

Figure 1: Number of Tweets

We also represent the impact of different media and personalities showing the total number of tweets and how many were reacted to by the audience.

The top contributors are analysed basing on the number of retweets that they received from their entire collection of tweets that they shared during the respective days. On pre-election day Nbs TV derived more engagement as it was providing a wider coverage of events leading up to the Election Day. The audience was more into knowing what is on ground and the media houses provided with live coverage on the events as they happened in real time. Newvisionwire an account related to New Vision a media house for the government registered less interaction as compared to the other independent media houses.

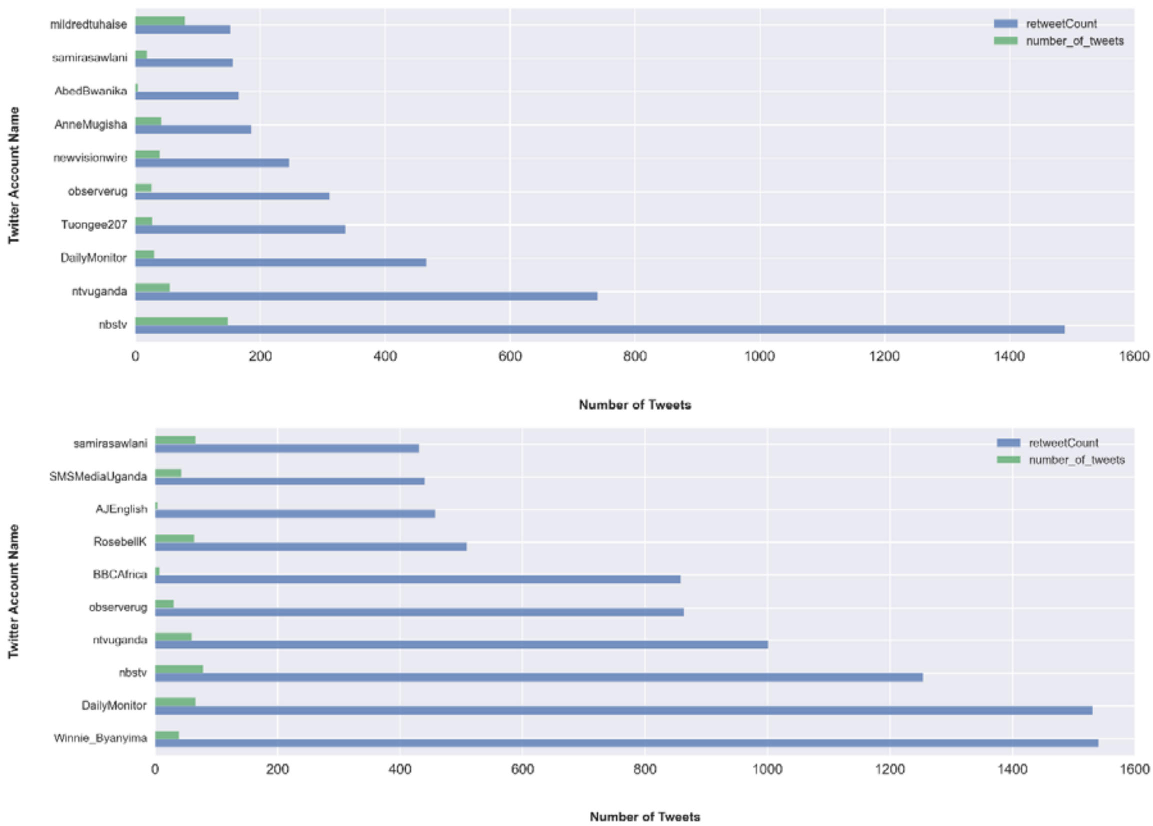

Figure 2: Top accounts and retweets

On the day of the elections Winnie Byanyima the wife to the main opposition candidate, Kizza Besigye, drove more engagement from social media as she posted about Uganda going to the polls. This trend showed the dominance of Winnie Byanyima considering that in the top 4 engagers of the Election Day she had the least number of tweets but with more retweets. The other top 6 engagers on the day of elections continued to drive engagement especially highlighted the peculiarities in the voting process including the delays of delivery of voting materials at some polling stations in the central region, which was considered the opposition stronghold.

Information Diffusion

We analyse how information flows from one user account and how it diffuses through the network and the impressions the data provides on actual events as they happen in real time. The tag cloud in Figure 3 shows a representation of the different topics that people were sharing on twitter during the pre-election and Election Day. The words with the highest frequency are represented with a higher font size compared to the less frequent ones. This shows the frequently used words alongside text in tweets that were posted. The major topic was that Ugandans were deciding their future for the next five years as they went to the polls, evident to the hashtag #UgandaDecides, it showed a spirit of patriotism amongst the twitter audience. Also Ugandans were not allowed to access social media websites and so most of the twitter audience were using Virtual Private Network (VPN) connections which hide one's location hence enabling them social media access. Due to some peculiarities in the voting process in some areas in the central region, including late delivery of voting materials, causing many to withdraw from polling stations whereas others using the twitter platform to encourage others to wait and vote, this resulted in some areas voting late into the night subsequently not including some of those results in the tallying of the national votes, there were some messages of disappointment to the Electoral Commission and the Chairman, Eng. Badru Kiggundu. Besigye the main opposition candidate and the incumbent Museveni, also appeared in tweets as they were considered the major rivals in the election process.

Figure 3: Tag Cloud with Most Used Hashtags

Analysis of sentiments

The most important part is the content of a tweet, the text, and that's where we look at to understand what people were talking about and perform some analysis. We perform sentiment analysis which is one of the interesting applications of text analytics. Although the term is often associated with sentiment classification of documents, in the broad sense it refers to the use of text analytics approaches applied to the set of problems related to identifying and extracting subjective material in text sources. In this work we use different tools which include the National Research Council Word-Emotion Association Lexicon (Mohammad, 2010). A data dictionary that has been built with a great number of English words with their associated score for eight different emotions and the two sentiments (positive/negative). Each individual word in the lexicon will have a Boolean "yes" (1) or Boolean "no" (0) for the emotions and sentiments, and thereafter we can calculate the total sentiment of a sentence by adding up the individual sentiments for each word in the sentence.

From the analysis we didn't have a significant change in trend of the sentiments, with the data presenting a higher bias towards positive sentiments, closely accompanied by trust, anticipation sentiments as depicted in Figure 4 and 5.

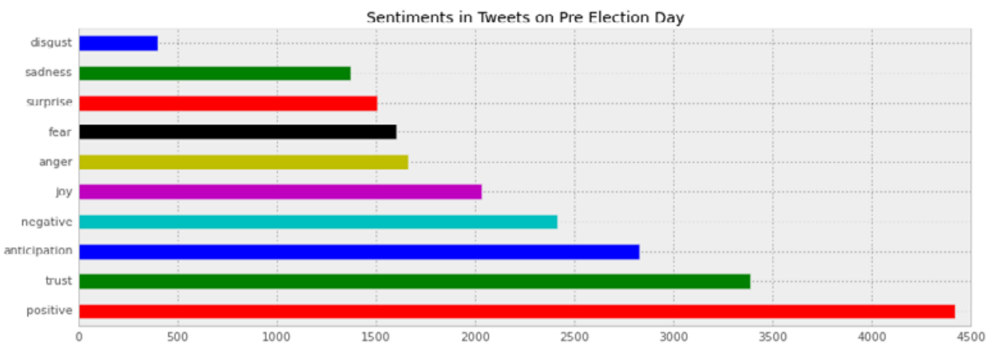

Figure 4: Sentiments in Tweets on Pre-election day

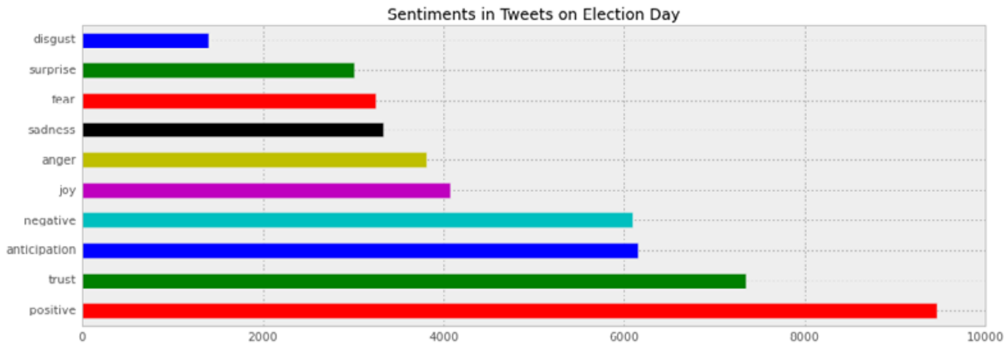

Figure 5: Sentiments in Tweets on Election Day

From the twitter data analysed there were more positive sentiments compared to negative ones for both the pre-election and election days as in Figure 4 and Figure 5, with emotions of trust and anticipation following closely. Emotions of disgust, surprise, sadness and fear were on the low end which could be attributed to a fair election exercise. The weights in the sentiments on the day of election was higher due to the higher number of tweets that were shared.

Semantic Orientation

After understanding the emotional polarities (sentiments) we perform another technique called semantic orientation. This technique is used in determining the relationship of words in text with positive and negative words (Turney, 2002) which is an easier test to perform since it's unsupervised, which means it doesn't require any labelled data for training.

The semantic orientation (SO) of a word is defined as the difference between its associations with positive and negative words. So we want to define the closeness of a word with positive and negative terms. The chosen measure of "closeness" is Pointwise Mutual Information (PMI) (Bouma, 2009), calculated as follows (x_1 $x_2$are terms):

$$PMI(x_1, x_2) = log \frac{P(x_1 \wedge x_2)}{P(x_1) . P(x_2)} \quad (1)$$

In Turney's paper, the SO of a word was calculated against excellent and poor, but of course we can extend the vocabulary of positive and negative terms. Using V+ for a vocabulary of positive terms and V- for the negative ones, the Semantic Orientation of a term t is hence defined as:

$$SO(x) = \sum_{\forall \epsilon V^+} PMI(x_1, x) - \sum_{\forall \epsilon V^-} PMI(x_1, x) \quad (2)$$

The Semantic Orientation of a term will have a positive (negative) value if the term is often associated with terms in the positive or negative vocabulary. The value will be zero for neutral terms i.e. the PMI's for positive and negative balance out, or more likely a term is never observed together with other terms in the positive or negative vocabularies. We use the opinion lexicon (Liu, 2012) which is a predefined set of positive and negative terms that will help us understand the semantic orientation, though it is limited to the English language.

Computing Term Probabilities

We compute P(x) (the probability of observing the term t) and $P(x_1 \wedge x_2)$ (the probability of observing the terms $x_1x_2$occurring together). Given the set of documents (tweets) D, we define the Document Frequency (DF) of a term as the number of documents where the term occurs. The same definition can be applied to co-concurrent terms. Hence, we can define our probabilities as:

$$D \vee$$

$$P(x) = \frac{DF_{(x)}}{} \qquad (3)$$

$$D \vee$$

$$P(x_1 \wedge x_2) = \frac{DF_{(x_1 \wedge x_2)}}{}$$

The document frequency for single terms was stored in their respective dictionaries while the document frequency for the co-occurrences was stored in the co-occurrence matrix.

Term co-occurrences

We were interested in the terms that occur together and the frequency, we want to discover the context so as to be able to give us much clearer insights about the understanding of the context whether positive and negative the words are being used, supporting the ideology of semantic similarity. So we understand the context of a term to the entire text contained in the tweet. We then build a co-occurrence matrix so that it contains the number of observations the two terms have been used together.

Computing the Semantic Orientation

And finally the semantic orientation of the two common words in the tweets collected and analysed

Pre-Election Day				Election Day			
Co-Occurrences	Frequency	SO	Normal	Co-Occurrences	Frequency	SO	Normal
Polling, Station	909	-44.572	-34.834	Polling, Station	1012	-63.451	-49.5885
Election, Uganda	577	-14.362	-11.224	Medium, Social	881	-220.401	-172.249
Phone, Polling	458	-43.427	-33.939	Material, Voting	366	-70.018	-54.7208
Ballot, Paper	434	-80.487	-62.903	Ballot, Paper	345	-222.513	-173.899
Mbabazi's, Support	404	13.757	10.751	Election, Uganda	297	-113.938	-89.0453
Ban, Phone	393	-48.186	-37.659	Station, Voting	281	-41.345	-32.3121
Election, Fair	382	-16.806	-13.134	Polling, Voting	261	-30.559	-23.8826
Polling, Voter	378	-22.836	-17.847	Arrested, Besigye	256	-257.574	-201.3
Phone, Station	378	-47.069	-36.786	Besigye, Kizza	239	-177.899	-139.032
Phone, Voter	369	-33.53	-26.205	Pm, Voting	230	-80.954	-63.2675
Museveni, President	260	-60.21	-47.056	Ballot, box	215	-221.003	-172.719
	Total SO	**-397.728**				**-1499.66**	
	Total Documents	**1073**				**1650**	
	Normals	**0.370669**				**0.908882**	
	Total	**1.279551**					

Table 3 shows the words that were seen together with the highest frequency and their respective semantic orientation, showing the nature of discussions that people were tweeting about frequently for instance, on the day before elections the Electoral commissioned issued a warning about people using their phones at the polling stations, which contributed to a number of people commenting on that statement and on the election day social media was shut down which led many twitter users to resort to Virtual Private Networks to access the internet, voting materials were also delivered late to some polling stations especially in some areas in the central region, with a number of stations starting to vote late in the afternoon and also the presidential candidate Kizza Besigye was also arrested on the same day sparking discussions on twitter on the matter.

To leave no room for chance we employed semantic orientation on the most frequent term co-occurrences. To compare the two data sets, the semantic orientations of the top 10 word co-occurences are normalised. It is discovered here that just because two words are popular it doesn't infer that it is necessarily in a positive or negative context, but to what extent are they in negative terms or positive.

A semantic orientation value of 0 indicates neutrality with a more negative value showing a more negative context whereas a more positive value indicating a more positive context. Most of the frequently used terms were identified to be used in a negative context with the strength in the negativity increasing on the day of the Elections.

Conclusion

On the Election Day, there was a drop in the number of social media users because social media sites like twitter, Facebook, YouTube had been shut down for security purposes. These sites could only be accessed by individuals who had knowledge about virtual private network softwares (software that facilitates creation of a secure connection to another network over the internet) that enabled them to access these restricted sites while shielding their browsing activities from prying eyes on public wireless fidelity (Wi-Fi). This affected us during the research work as we were able to make limited inferences because of the limited data.

Most twitter users keep their locations unknown and thus disable their Global Positioning Systems (GPS), so that their geo locations remain anonymous, this limited our analysis to the available information and thus we were not able to do geographical sentiment analysis.

From the sentiments and emotions discovered on twitter, showed a generally positive mood during the elections, the semantic orientations from the most frequent terms in the tweets collected showed a negative attitude towards the way the electoral process was handled, from the banning of mobile phones, switching off social media access with no clear reason, delay of delivery of voting materials at some stations in the central region resulting into late voting and frustration of voters and the arrest of the main opposition candidate Kizza Besigye.

Future work

In this work we used lexicons and dictionaries based on the English language, which was difficult to apply to the twitter data set that we analysed. There was need to customize the dictionaries to the local context of the audience involved in this research. The data had a lot of synonyms in the local languages, which were not recognised hence dropping some words and meanings which could have contributed to a much finer analysis when we applied the corpora of tweets to this process explained in this work. In future we seek to build more customized lexicons and word dictionaries to incorporate the way majority of Ugandans communicate on social media to be able to enhance the accuracy of our results.

References

Achrekar, H. a.-H. (2011). *Predicting flu trends using twitter data. Computer Communications Workshop*s (INFOCOM WKSHPS), 2011 IEEE Conference (pp. 702-707). IEEE.

Asur, S. a. (2010). *Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology* (WI-IAT). Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference. 1, pp. 492-499. IEEE.

Bermingham, A. a. (2011). *On using Twitter to monitor political sentiment and predict election results.*

Bouma, G. (2009). *Normalized (pointwise) mutual information in collocation extraction*. Proceedings of GSCL, 31-40.

Buche, A. C. (2013). *Opinion mining and analysis: A survey*., (p. 10).

Cipesa. (2016). *Analysis of Twitter Activity During the 2016 Presidential Debates in Uganda, Monitoring Uganda Elections Series 01, #UgDebate16*. Kampala: Cipesa. Retrieved May 23, 2016, from http://www.cipesa.org/?wpfb_dl=210

Cipesa. (2016). *Ugandans Turn to Proxies, VPN in Face of Social Media Shutdown*. Kampala: Cipesa. Retrieved May 23, 2016, from http://cipesa.org/2016/02/ugandans-turn-to-proxies-vpn-in-face-of-social-media-shutdown/

Jahanbakhsh, K., & Moon, Y. (2014). The predictive power of social media: on the predictability of US presidential elections using Twitter. *arXiv preprint arXiv*:1407.0622.

Lei, L. B. (2012). *Mining text data: A survey of opinion mining and sentiment analysi*s. Springer.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining: Synthesis Lectures on Human Language Technologies, vol. 16.* Morgan & Claypool Publishers, San Rafael.

Liu, B. a. (2012). *Mining text data: A survey of opinion mining and sentiment analysis*. Springer.

Mohammad, S. M. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text (pp. 26-34). Association for Computational Linguistics.*

Musisi, F. (2016, Feb 14). Queries as Electoral Commission bans phones at polling stations. *Daily Monitor*. Retrieved from http://www.monitor.co.ug/News/National/Queries-as-Electoral-Commission-bans-phones-at-polling-stations/-/688334/3075780/-/6yv2rr/-/index.html

URN. (2016, Feb 13). EC apologises for 20,000 'ghosts' on voters register. *The Observer*. Retrieved from http://observer.ug/news-headlines/42596-ec-apologise-for-20-000-ghosts-on-voters-registers

Kiyonga, D. and Lubwama, S. (2016, May 23). Kayihura ordered social media shutdown- UCC. Kampala. *The Observer. Retrieved from http://observer.ug/news-headlines/43165-kayihura-ordered-social-media-shutdown-ucc*

Ritter, A. a. (2011). Named entity recognition in tweets: an experimental study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 1524-1534)*. Association for Computational Linguistics.

Ritterman, J. O. (2009). Using prediction markets and Twitter to predict a swine flu pandemic. *1st international workshop on mining social media Vol. 9*. Retrieved from http://homepages.inf.ed.ac.uk/miles/papers/swine09.pdf, (pp. 9-17).

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media, Inc.

Tumasjan, A. a. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 178-185.

Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics.*

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 115-120). Association for Computational Linguistics.