VOL. 10, NO. 3, FEBRUARY 2015 ISSN 1819-6608

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

OPINIONS FROM TWEETS AS GOOD INDICATORS OF LEADERSHIP AND FOLLOWERSHIP STATUS

Salamatu Osanga I. and Naomie Salim
Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
E-mail: drnaomiesalim@gmail.com

ABSTRACT

Scores of public opinion about two popular world leaders collected from tweets based on the sentiment they exhibited were classified using two Machine learning techniques (Naïve Bayes and Support vector machines), and four features (Words, unigrams, bigrams and negation) for the classification, we found that the Naïve bayes with unigram features attained a high accuracy of up to 90% therefore indicating that tweets can be used to suggest potential candidates in political election and ways to improve a leaders reputation.

Keywords: Keybased lexicon, subjectivity, opinion mining, twitter, leadership, followership, evaluation.

INTRODUCTION

In order to know the certain number of people for or against a leader, the simplest thing to do is to take a random sample obtained through a survey or an election. Survey and polling methodology, extensively developed through the 20th century gives numerous tools and techniques to accomplish representative public opinion measurement (O'Connor et al, 2010).

The sudden eruption of text-based facilities on social media has enable citizens or followers to easily air out their views and beliefs. These views and beliefs also termed as Opinion could range from making reviews of products and services to expressing opinions on topics like health, education, tourism and even sensitive topics like politics. As a societal norm, the idea of leadership must evolve and keep pace with all societal changes as they occur. As our society changes at a rapid pace our understanding of leadership must change along with societal needs, else it becomes irrelevant and obsolete. Also the free availability of social media has thus occurred as a means of responding directly to the surge of interest that deals with opinions and use of information technologies to seek out and understand the opinions of others.

A vital question here is can publicly available data be analysed in order to infer population attitudes in the same manner that public opinion pollsters query a population? If so, then mining public opinion from freely available text content could be a faster and less expensive alternative to traditional surveys and polls. Extracting public opinion from social media is not only hectic and tedious but also challenging as well. Hence exploration of the rich context of this unstructured data has promoted the alliance of natural language computational models and computational linguistics research.

In this paper, we analyse and classify public opinion about two world known leaders collected from the popular microblogging site Twitter. The choice of these leaders fell on Barack Obama, the United States president and Nelson Mandela a former South African president and an anti-apertied hero. In selecting theses mentioned leaders, the following factors were highly considered:

a) These leaders had acquired fame worldwide, either in a positive or negative way
b) These leaders are timeless and unforgettable celebrities, as they always remembered and made reference to in societies even in death (Mandela)
c) The ease of data availability expressed on the selected leaders.
d) The year they each led. As they belong to the older and younger generation.

First we generated words and build our corpus, next we performed a linguistic analysis of our corpus and showed how to build a sentiment classifier that uses the collected corpus as a training data

RELATED WORK

The dramatic rise of text-based social media has made opinion mining and sentiment analysis a field of surging interest for many researches. A broader view of the existing approaches and techniques were presented in (Pang and Lee, 2008). In their survey, the authors describe existing techniques and approaches for an opinion-oriented information retrieval. However, not many researches in opinion mining considered blogs and even much less addressed microblogging. In (O'Connor et al., 2010), the authors use twitter to construct a corpora for sentiment analysis and further used unigram features (keywords related to polls) as indicators of public opinion.. The authors applied SVM and NB algorithms to classify opinions and then investigated Factors affecting the classifier's performance. The Authors were able to obtain an accuracy level of up 90%. Tumasjan in (Tumasjan et al., 2010) used Party names to collect data from twitter to further f form training set for sentiment classification. SVM and Naive Bayes were able to obtain up to 80% of accuracy level on the collected tweets. In

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

(Go et al., 2009), authors used Twitter to collect training data and then to perform a sentiment search. The authors construct corpora by using emoticons to obtain "positive" and "negative" samples, and then use various classifiers. The best result was obtained by the Naive Bayes classifier with a mutual information measure for feature selection. The authors were able to obtain up to 81% of accuracy on their test set. However, the method showed a bad performance with three classes ("negative", "positive" and "neutral").

DATA

We begin by discussing the data used in this study: based on a list of 57 items of prototypical leadership attributes generated by test subjects, Offermann et al. (1994) were able to generate an eight dimensional factor structure vis a vis : 'Dedication' (e.g. commitment, devotion); 'Sensitivity' (Consciousness and understanding); 'Tyranny' (persuasive, compelling); 'Charisma' (inspiring and alluring); 'Attractiveness' (well-groomed); 'Masculinity' (firmness and virility); 'Intelligence' (intellect, knowledgeable); 'Strength' (courage and stability). These are further called collective expectations or 'impression dimensions'. They function as cognitive points of orientation in the perception of leadership, and on the other as social dramatization or social reversion options. Tweets are collected based on these leadership attributes

Twitter corpus collection

Data was collected using twitter API. Twitter provides two APIs namely: REST and Streaming API's. The main difference between these two API's is that Streaming API supports long-lived connection and provides data in almost real-time while the REST APIs support short-lived connections and are rate-limited (i.e. only certain amount of data are available for download in a day) while the Streaming API enables access to currently trending issues on twitter. For the purpose of this research the streaming API was used for the tweets collection.

First, tweets are collected via the API using advanced search (Go et al., 2009) based on the keywords "Obama's leadership" and "Mandela's leadership". Next, a bootstrapping module that aids in streaming tweets called tweepy2 was incorporated into the API application and used to collect more tweets based on the assumption that some tweets will contain some sampled words about the leaders attributes but will not contain the names of these leaders. These assumptions are made on some pre-defined decisions made beforehand and a well-defined set of features as the seed list. So the search was refined to include the keywords "Obama (dedication, sensitivity, tyranny, intelligence, charisma, attractiveness, masculinity, strength)" and "Mandela (dedication, sensitivity, tyranny, intelligence, charisma, attractiveness, masculinity, strength)". In our research, we use English language. However, our method can be adapted easily to other languages since Twitter API allows specifying the language of the retrieved posts.

Corpus analysis

From the tweets, we are interested in assessing people's aggregate opinion on two leaders. In doing this; the task can be broken down into two sub problems:

- Opinion retrieval: identify messages relating to the topic.
- Opinion estimation: determine whether these messages express positive or negative opinions about these leaders.

Owing to enough data at our disposal, the work was formulated as a topic-sentiment model (Mei et al. 2007), in which the topics and sentiment of documents are jointly inferred. We therefore opt to use a transparent, deterministic approach based on prior linguistic knowledge, counting instances of positive-sentiment and negative-sentiment words in the context of a topic keyword.

Opinion retrieval

We only use opinions containing a topic keyword, manually specified for each tweet. The activities used are specified as follows;

Subjectivity detection

Three people were assigned the task of rating and determining the subjectivity of each given tweet based on the basic constituents that forms a subjective sentence as shown by Kim and Hovy (2004) and Wiebe et al (2004). In general, an opinion expression is made up of an opinion holder, an opinion indicator, an opinion object or object feature, and one or more polar words for expressing sentiment orientation (liu, 2010). The opinion holder often expresses his or her opinion on a target via some special verbs such as "(looks at)" and "(think)" . Such verbs are often termed opinion indicators. For example, taking a look at this opinionated sentence, we can deduce that; "(She did accuse the Nigerian government and military for lying irresponsibly.). " (She)" is an opinion holder that expresses the opinion, "(accused)" as an opinion indicator, "(the Nigerian government, military)" are the opinion target of. "(Irresponsibly)" and "(lie)" are two polar words that express a negative orientation. Motivated by the above observations, three types of lexical cues for subjectivity indication will be considered. They are;

- Name entities or pronouns such as leadership, leader he etc.
- Opinion polar words such as good, bad, courageous e.tc
- Opinion target i.e. Obama and Mandela

VOL. 10, NO. 3, FEBRUARY 2015 ISSN 1819-6608

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Based on the above subjectivity lexical cues, these raters will help detect the subjectivity of the sampled tweets. The tweets were manually assessed for annotation based on the Gold standard for manual annotation (wiebe et al, 2004),

Feature extraction

In order to extract leadership related features, certain features have to be selected from the entire annotated corpus. Selection here is based on the frequency of words known as the information gain criterion. The collected dataset is used to extract features that will be used to train our sentiment classifier. Pang et al. have obtained better results by using a term presence rather than its frequency (Pang et al., 2002). We have experimented with words, unigrams, bigrams, and negations. Pang et al. (Pang et al., 2002) reported that unigrams outperform bigrams when performing the sentiment classification of movie reviews, and Dave et al. (Dave et al., 2003) have obtained contrary results: bigrams and trigrams worked better for the product-review polarity classification. We tried to determine the best settings for the micro blogging data. On one hand high bigrams, should better capture patterns of sentiments expressions. On the other hand, unigrams should provide a good coverage of the data. The processes of obtaining these features from Tweets are as follows:

- **Tokenization:** in this step, each tweet is divided into smaller units known as tokens. These tokens consist of words.
- **Normalization:** the tokens are further expanded to give more consistent elements with the same textual form. This typically involves the use of rule-based text processing. Examples include;
  - 12/12/1984! – 12 December 1984
  - Pres. - President
  - Gooooood - Good
  - Converting all words to lower case
  - Removing repeated characters
- **Stemming:** each individual term is reduced to its stem using the Porter stemmer's algorithm.

Stop words removal: wordlist was also used in order to get rid of the features that don't convey meaningful sentiment. Example of such words includes; are, us, is, he etc.

Keywords lexicon generation: a detailed literature review is done also on leadership traits and attributes of which a total of 8 attributes are handpicked as recommended by known researchers in the field of leadership. These attributes are form list of Offer man et al (1994), 8 dimension prototypical leadership attributes generated by text subjects. . The 8 attributes are;

1. Dedication
2. Sensitivity
3. Tyranny
4. Charisma
5. Attractiveness
6. Masculinity
7. Intelligence
8. Strength

These attributes are considered as the seed list and are further use in generating leadership oriented words about leaders. The process of generating these words involves building two discriminatory-word lexicons based on polarities. The first word- lexicon contains words indicating positive sentiment while the second word-lexicon contains words indicating negative sentiment. The seed list was expanded using Word Net where each attribute synonym is considered as a positive word whereas attribute antonyms are considered negative words simultaneously (Kim and Hovy, 2005). The negative-keyword lexicon contains 131 words and the positive-keyword lexicons contain 325 words.

Constructing unigrams and bigrams: We made a set of unigrams and bigrams out of consecutive words. First we input all words as a sentence, next Place a window on each individual word and split to give a separate word (unigram) or split on first two words (bigrams). Continue until all the words are exhausted (Wilson et al., 2005). Table-1 depicts the list of constructed unigram and bigram features.

Table-1. A list of constructed unigram and bigram features.

	Unigram Features	Bigram features
1	Courage	courage + freedom
2	Freedom	Freedom + charisma
3	Charisma	charisma + wisdom
4	Wisdom	wisdom + charm
5	Charm	Charm+ tolerant
6	Tolerant	tolerant + sensible
7	Sensible	sensible + extravagant
8	Extravagant	extravagant + courage

Constructing negations: A negation word such as no, not, and, never and also some words that follow patterns such as "stop" , "quit" and "cease" change the orientation of opinion words in the following way (Wilson et al, 2005):

I. Negation + Negative word = Positive opinion
II. Negation + Positive word = Negative opinion

Opinion estimation: A Tweet is defined as positive if it contains any positive word and negative if it contains any negative word. (This allows for messages to be both positive and negative.), since Twitter posts are so short (about 11 words). We build a sentiment classifier using the Naive Bayes classifier and SVM (Alpaydin, 2004). However the Na¨ıve Bayes classifier yielded the best

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

results. Na¨ıve Bayes classifier is based on Bayes' theorem (Anthony J, 2007).

$$P(s|M) = \frac{P(s) \cdot P(M|s)}{P(M)} \quad (1)$$

Where s is a sentiment, M is a Twitter message. Because, we have equal sets of positive, negative and neutral messages, we simplify the equation:

$$P(s|M) = \frac{P(M|s)}{P(M)} \quad (2)$$

$$P(s|M) \sim P(M|s) \quad (3)$$

We trained the two classifiers using different features: words, unigrams, bigrams and negation
The training was carried out based on

- **Training:** this entails learning a model based on the prediction made on the test data. During the training, the presence of each feature value in each of the classes (negative, positive) is counted as shown in Table-2.

Table-2. Features observed probabilities.

Words	P(Pos)	P(Neg)
Leaders	4/6	2/6
Honesty	3/8	5/8
Tyranny	1/3	2/3
Intelligence	1/2	1/2
Charisma	1/2	1/2

Furthermore, the observed probabilities of features in each class (negative, positive) are computed. This count is about how many of the instances are positive and negative as shown in Table-3.

Table-3. Class probabilities.

L(Pos)	L(Neg)
2/20	3/20

- **Prediction:** in determining the sentiment of the new tweet, 2 hypotheses have to be tested. These are:
  positive and negative. First a learning algorithm will look for the presence of any of the selected features in the new tweet. Next, the probabilities of each of the hypothesis are computed from the found feature using the Machine leaning algorithms. For example, in order to ascertain the probability of a tweet belonging to either positive or negative hypotheses, given an evidence First lets us P(pos/E) and P(neg/E) to denote the probability of positive and negative hypothesis respectively, where E represents the evidence.

Next let's compute the probabilities of the two hypotheses in Naive Bayes using the probabilities of the words 'honesty'; and 'charisma' obtained from the learned model above. Substituting them in the formulas for each class gives;

P(pos/E) = $(Ppos_{honesty} * Ppos_{charisma}))$ * L(pos)
$Ppos_{honesty}$ and $Ppos_{charisma}$ are the probabilities of 'honesty' and 'charisma' in the positive class. and L(pos) is the probability of the positive class. These values are from the above learned model. $Ppos_{honesty}$ = 3/8, $Ppos_{charisma}$ = 1/2 and L(Pos) = 2/20. With these values the probabilities becomes;
P(pos/E) = (3/8 * 1/2) * 2/20
P(pos/E) = 6/320
In the same manner,
P(neg/E) = $(Pneg_{honesty}$ * $Pneg_{charisma}))$ * L(neg)
P(neg/E) = (1/2 * 1/2) * 3/20
P(neg/E) = 3/40

Now that the probability of Twitter post being generated by each of the sentiment classes is known, we can decide what its sentiment is. Clearly, this tweet is positive as the P(pos/E) has the highest value. This is essentially how Naive Bayes works.

METHODOLOGY

We have tested our classifier on a set of real manually annotated Twitter posts. We used the same evaluation set as in (Goet al., 2009). The characteristics of the dataset are presented in Table-4

Table-4. The characteristics of the evaluation dataset.

Polarity	Number of samples
Positive	800
Negative	800
Total	1600

We compute accuracy (Manning and Sch¨utze, 1999) of the classifier on the whole evaluation dataset, i.e.:

Accuracy: This measures the proportion of tweets that are correctly obtained in the corpus. The percentage of the correctly classified objects use in calculating the accuracy of a classifier is calculated as follows:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Precision: the class precision defines the probability that if a random tweet is classified with this class, then it is considered the correct choice i.e.

ARPN Journal of Engineering and Applied Sciences

Precision measures the exactness of a selected target in a given corpus.

Precision for the positive class for instance is calculated as follows:

$$P = \frac{TP}{TP+FP} \tag{5}$$

Recall: this measures the completeness which is the proportion of targeted corpus that the system selects. The class recall defines the probability that if a random tweet should be classified with this class, then the decision is complete. Recall for the positive class for instance is calculated as follows:

$$R = \frac{TP}{TP+FN} \tag{6}$$

Where

TP (true positives) denotes the number of positively-labelled test tweets that were correctly classified as positive;
TN (true negatives) denotes the number of negatively-labelled test tweets that were correctly classified as negative;
FP (false positives) denotes the number of negatively-labelled test tweets that were incorrectly classified as positive;
FN (false negatives) denotes the number of positively-labelled test tweets that were incorrectly classified as negative.

RESULTS

We tested the impact of the features on the classifiers performance. The results of this comparison are presented in Figure-1 to 3. As we see from the Table-5, the best result is boldfaced. The best performance is achieved when using unigrams. This may be due to the ability of unigrams feature in capturing expressive sentiment.

Table-5. Comparison of the models performance.

Models	NB				SVM			
features	Words	**Unigrams**	Bigram	Negation	Words	Unigrams	Bigram	Negation
Accuracy (%)	86.3	**91.4**	89.8	84.5	83.9	84.1	82.9	80.9
Precision	0.86	**0.91**	0.89	0.844	0.83	0.85	0.82	0.81
Recall	0.85	**0.91**	0.88	0.84	0.82	0.84	0.82	0.81

The highest accuracy achieved for opinion polarity classification is 91.41% achieved by the NB + unigram feature. A closer look at the feature shows competition between the unigram and bigram features. All the features performed above average and each achieved its highest accuracy rate when about 70 to 80% of the dataset was used for training.

As a result, the NB model has more exactness and completeness in classifying that part of retrieved tweets that are relevant and can also help pinpoint a particular class having difficulty with prediction better than SVM. Hence, NB has again outperformed the SVM model in exactly and precisely predicting and classifying words about leaders into classes of positive and negative using feature frequency representation.

Correlation analysis: Is text sentiment a leading indicator of leadership and followership status? Judging by the performance of our modelled classifiers and followers reaction to the challenge of leadership exhibited and tagged as the "Twitter or Social media revolution" that lead to the uprisings in North America and some part of middle east. It is apparent that the sentiment ratio captures the broad trends in text.

CONCLUSIONS

In the paper we find that a relatively simple sentiment detector based on Twitter data replicates follower's confidence on leadership approval. While the results do not come without caution, it is encouraging that expensive time-intensive survey can be supplemented or supplanted with the simple-to-gather text data that is generated from online social networking. The results suggest that more advanced NLP techniques to improve opinion estimation may be very useful. The textual analysis could be substantially improved. Besides the clear need for a more well-suited lexicon, the modes of communication should be considered. Many techniques from traditional survey methodology can also be used again for automatic opinion measurement. For example, polls routinely use stratified sampling and weighted designs to ask questions of a representative sample of the population. Given that many social media sites include user demographic information, such a design is a sensible next step. Eventually, we see this research progressing to align with the more general goal of query-driven sentiment analysis where one can ask more varied questions of what people are thinking based on text they are already writing. Modelling traditional survey data is a useful application of sentiment analysis. But it is also a stepping stone toward larger and more sophisticated applications

ACKNOWLEDGEMENTS

This work is supported by the ministry of higher education (MOHE) and Research Management Centre (RMC) at University Technology Malaysia (UTM) under research University grant category (Vot: Q.J130000.2528.07H89). We would also like to thank the reviewers for the helpful criticism and suggestions.

REFERENCES

O'Connor B., Balasubramanyan R., Routledge B.R and Smith, N.A (2010). "From tweets to polls: Linking text

ARPN Journal of Engineering and Applied Sciences

sentiment to public opinion time series," Proceedings. Of the International AAAI Conference on Weblogs and Social Media, pp. 122-129.

Pang B., Lee L. and Vaithyanathan, S. (2002) Thumbs up? sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10, EMNLP '02, pages 79–86. Association for Computational Linguistics.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Found. Trends Inf. Retr., 2(1-2):1-135.

Bo Pang., Lillian Lee. and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the Conference.
on Empirical Methods in Natural Language Processing(EMNLP), pages 79-86

Tumasjan A., Sprenger T. O., Sandner P. G. and Welpe I. M. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. Fourth International AAAI Conference on Weblogs and Social Media.

Go A. Huang G. and Bhayani R.. 2009. Twitter sentiment analysis. Final Projects from CS224N for Spring 2008/2009 at The Stanford Natural Language Processing Group.

Offermann L.R., Kennedy J.K. and Wirtz, P.W. (1994). Implicit leadership theories: Content, structure, and generalizability. Leadership Quarterly, 5, 43-58.

Mei Q., Ling X., Wondra M., Su H. and Zhai C. X. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. Proceedings of the 16th International conference on World Wide Web.

Kim S. and Hovy E. (2004). Determining the sentiment of opinions. Proceedings of the 20th international conference on Computational Linguistics, COLING '04. Association for Computational Linguistics.

Liu Bing. (2010) Sentiment Analysis and Subjectivity, in Handbook of NaturalLanguage Processing, Second Edition, N. Indurkhya and F.J. Damerau, Editors.

Manning C.D., Raghavan P. and. Schtze H. (2009) An Introduction to Information Retrieval. Cambridge University Press.

Wiebe jaynce, Wilson Theresa , Bruce Rebbeca, Matthew Bell, and Melanie Martin. 2004. Learning Subjective Language. Computational Linguistics, 30(3):277-308.

Dave K., Lawrence S. and Pennock D.M (2003) Mining the peanut gallery: opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web, WWW '03, pages 519-528. ACM.

Kim S. and Hovy E. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. Proceedings of ACL/Coling Workshop on Sentiment and Subjectivity in Text, pp1-8

Wilson T., Wiebe J. and Hoffmann P. (2005) Recognizing contextual polarity in phraselevel sentiment analysis. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05, pages 347–354. Association for Computational Linguistics.

Alpaydin E. 2004. Introduction to Machine Learning (Adaptive Computation and Machine Learning). The MIT Press.

Anthony H. J. 2007. Probability and Statistics for Engineers and Scientists. Duxbury, Belmont, CA, USA., pp.1-8.