POLITICAL MINER: OPINION EXTRACTION FROM USER GENERATED POLITICAL REVIEWS

Muhammad Zubair Asghar[1], RahmanUllah[1], Bashir Ahmad[1], Aurangzeb Khan[2], Shakeel Ahmad[1], Irfan Ullah Nawaz[1]

[1]Institute of Computing and Information Technology,[2]Institute of Engineering and Information Technology, Faculty of Sciences, University of Science and Technology, Bannu

Emails:

zubair@gu.edu.pk,rahman_lakki@yahoo.com,bashahmad@gu.edu.pk,aurangzeb.ustb@gmail.com,shakeel@gu.edu.pk, irfannawaz88@yahoo.com

Phones:0333-9968518, 0345-8981401, 0322-9527806, 0333-9760222, 0321-9606033, 0302-2959588

ABSTRACT: *Millions of users share their opinions on micro blogging websites through short messages on different topics. Opinion mining determines the mood, feelings, emotions or opinions of other people about services, politics, products and policies.In this paper we introduce a rule based opinion mining system that will evaluate the opinion of users they shared on twitter in response to some key political discourse. This framework will automatically crawl the tweets from twitter API, categorize text into subjective and objective tweets, calculates the polarity score at tweet and word level, and generate an aggregate of more than 500 political tweets. This approach shows impressive results and out performs the baseline method. The effectiveness of the proposed system is presented in a case study.*

Keywords: ?

INTRODUCTION

The growing use of World Wide Web facilitates people by providing great chances of collaboration and sharing of resources. Attractive nature of different social networking sites such as Twitter[1], YouTube[2], Facebook[3]etc., have fascinated many people. Social networking sites like Facebook facilitate communication within personal circle such as friends etc., whereas microblogging sites allow users to make new postings openly as well as enable them to comment publicly on already published issues. In order to increase political contribution among voters, microblogging sites like Twitter play vital role [2]. Twitter provides free service for the users and people share their ideas in the form of tweets. In addition it provides a lot of other services for the users such as following other users, key word searching in the tweets, retweeting facility, and favorite the comments of others and so on. It is evident from the previous research that in the months of election social media use is extensively increased and statistics showed that 22% of adults are in the queue for political conversation [1]. The most eminent example of such studies include of Brak Obama who effectively utilize social media networking sites within his last election campaign [3].

Since Twitter service is being used effectively for conversation as well as for collaboration [4] and that tweets represent electronic word-of-mouth communication [5]; it is possible from these discussions that we can observe the political moods and feelings by interception. General characteristics of Twitter, such as daily gossips, sharing facts, news reporting and chat, can include political views and sentiments. This can be easily observed in the days of elections when people talk about political parties, their leaders and political issues more often. Despite of the fact that Twitter may prove great opinion producing channel, it may face different type of challenges in the analysis of these opinions. For this purpose opinion mining tools are required in order to make full use of these user generated opinions.

[1]www.twitter.com

[2]www.youtube.com

[3]www. facebook.com

Opinion mining tries to detect the opinions, feeling or emotions expressed in the form of text [6]. Political opinion mining tries to detect political opinions or attitude given in the text. Currently many research studies have been conducted on applying sentiment analysis techniques to twitter data for extracting user opinion about political issues [7]. However, very little research is done to classify this extracted information to positive, negative or neutral tweets.

This study aims to devise a rule based subjectivity classifier, capable of mining user tweets shared on twitter during some key political event. In this paper we present a framework for subjectivity and objectivity classificationwhich will becompatible with both annotated and un-annotated dataset. We collect user tweets from twitter API, calculate its polarity score at word and sentence level and classify the tweets as negative and positive. Our dataset comprises of more than 500 tweets discussing current election in Pakistan. This paper is organized as follows: The related work is presented in section 2. Section 3 describes the methodology. In section 4 experimental results are presented. The final section concludes with discussion on future directions.

RELATED WORK

In this section we will discuss sentiment analysis of twitter messages regarding political domain.

Research on user reviews, posted on micro blogging networking sites, associated to political issues is very recent. In this regard Twitter remains a priority for people interested in politics. Kim [8] examined tweets during Korean Elections for year 2010 and concluded that during elections people used twitter service for three main reasons: for the sake of entertainment, for seeking political information and for social benefits. During the election of US 2010 researchersdetermined that, twitter proves to be a great platform for establishing political environment in numerous countries: providing a rich source of information regarding political views [9].

Twitter is a flexible communication channel as it can focus on the particular topic during some specific event as well as it represents other sort of communication during that event. Since Twitter service is being used effectively for conversation as well as for collaboration [4].Peoplemake new postings and also comment publicly on already published issues. In these commentspeople use opinion words (positive, negative or neutral) to dress their views.It is possible from these discussions that we can observe the political moods and feelings by interception.

Many approaches have been adopted for performing sentiment analysis on social media sites. Knowledge based approaches classify the sentiments through dictionaries defining the sentiment polarity of words and linguistic patterns [19], [20].Specifically, for twitter sentiment analysis not a single approach has been used by researchers. In this regard hybrid approach has been adopted by combining methods based on lexicon with those based on machine learning and natural language processing techniques, in order to get advantage of both content as well as connectivity patterns among twitter users.

The current work on Twitter sentiment analysis is based on constructing and using dictionaries for extracting polarity score at word level. For labeling the polarity of tweets O'Connor et al [10] used subjectivity lexicon: OpinionFinder. Further they implemented temporal smoothing technique for capturing the overall sentiment score. Bollen et al [11] performed sentiment analysis of mood expressed publically on twitter by using Profile of Mood States: a well-established tool for measuring semantic expressions of sentiment given in the text. Moreover, Tumasjan et al [12] showed that the messages posted on micro blog sites are the effective way of expressing political moods and opinions. Their analysis is based on Linguistic Inquiry and Word Count: software for text analysis.

Jonathon [13] classifies tweets based on the emoticons given by the user while expressing their sentiments and opinions on twitter. Other studies in this area focus on user's network. For example Tan et al [14] exploit relationship information from user network, in their work for polarity associations of user reviews.

However, our approach to mine the political tweets corresponding with elections is to provide a better and more effective way for opinion mining.

METHODOLOGY

Our approach to find semantic orientation of the opinion words present in political tweets is a hybrid one, using corpus based as well as dictionary based techniques. Features like emotion icons and capitalization of words are also under consideration as they are included in the intensification of the tweet and largely appear in the informal social media language. The overall data flow of the proposed system is given in figure 1 and the working is abstracted in algorithm x.

Pre-processing

In our approach preprocessing proceeds in the following way.All URLs (WWW.example.com), hash tags (#topic), targets (@username), Twitter special Words ("e.g. RT") are removed. Preprocessing module calculates the fraction of the words in Caps except nouns. Spell correction module is introduced and repeated character is tagged by a weight previously defined. This is done to highlight the normal and intensified words.The sentiment polarity of emoticons is annotated manually and their scores are obtained from the table.The occurrence of exclamation marks is counted and remaining punctuation marks are removed. POS tagger is used to tag verbs, adjectives and adverbs [16].

Semantic Score for Adjectives, Adverbs and Verbs

In this section, we calculate the semantic score of the opinionative terms: adjectives, verbs and adverbs.Our approach uses dictionary based method to get semantic orientation of verbs and adverbsand corpus based method to get the semantic orientation of adjectives. An Adjective describes a noun and qualifies objects. Since semantic orientation of adjective is domain dependent, thereforewe implement corpus based methods to manipulate it in Twitter domain. We adopt the work performed by [17] and apply a log-linear regression model with linear predictor(equation 1)

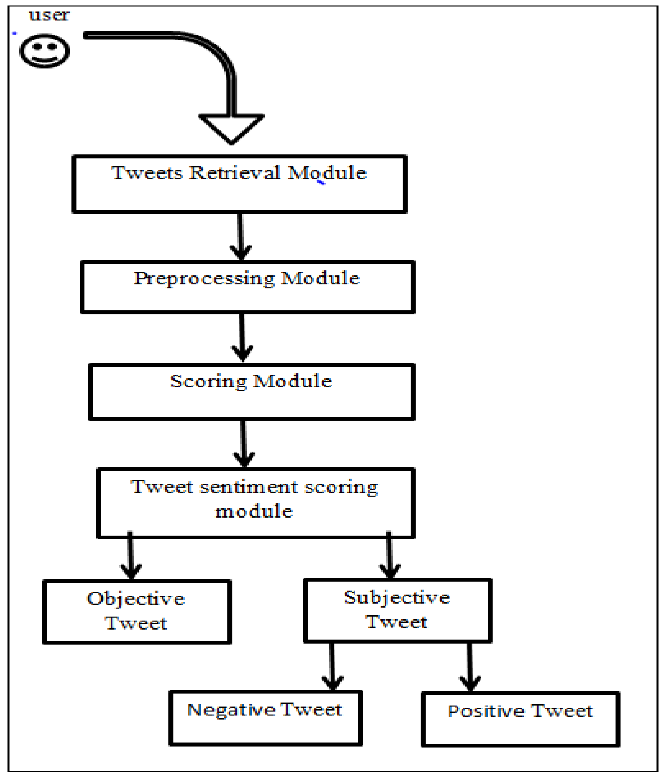

Fig. 1. Architectural representation of the proposed system

to calculate the semantic orientation of adjectives.

$$\eta = W\,T\,x \qquad (1)$$

In equation 1, "x" determines the vector of observed counts in the various conjunction categories for a specific adjective pair and "w" represents a weight vector learnt during training. The response y is non-linearly related to η through the inverse logit function

$$y = \frac{e^{\eta}}{1+e^{\eta}} \qquad (2)$$

In equation 2, the value of "y" mentions correlation between words. Initially, seed list of adjectives was assigned semantic scores manually. Manually assigned values and similarity value "y" is used to compute the conjoined semantic score of adjectives.

Verbs are also sentiment carriers and play important role in tweet sentiment. We use dictionary based methods for the semantic orientation of verbs and adverbs as they are domain free. Initial seed list containing positive and negative

Sentiment score of verbs and adverbsis extended by using Wordnet [18]. Further commonly used verbs and adverbs are manually annotated and values ranging from -1 to +1 are assigned.Adverb strength helps in assessing whether a document gives a perfect positive opinion, strong positive opinion, a slight positive opinion or a less positive opinion. For example; one user says, “This is a very good politician” and;other says, “This is a good politician”.

Algorithm1:FindingSubjectivity and Sentiment Score of Tweets

Input: SWN = <W, pos, pol_score>: Lexicon
T Tweet
Thr: Threshold
Negation List = {not, never…}
Context-Shift-List = {but, however…}
Enhancer-Reducer-List = {slightly, very…}
N (op) presents the number opinion groups and emoticons in the tweet.
Wc mention the words fractions capitalized in the tweet.
Lr show the count of repeated letters
En show the count of exclamation marks present in the tweet
W(AGi) denotes score of the ith adjective group,
W(VGi) denotes the score of the ith verb group,
W(Ei) denotes the score of the ith emoticon,
Nei denotes the count of the ith emoticon,
Output: Word sentiment score,
Tweet sentiment score,
Objective,
Subjective (positive, negative) Tweets
Begin:
1. Get (W, POS, largest_sent_score) from SWN;
2. For tweet T calculate (Wc, Lr, En);
3. Compute sentiment Sint(T) for intensifiers

$$\mathrm{Sint(T)} = \frac{\left(1+\frac{Wc+Lr+En}{3}\right)}{N_{(op)}}$$

4. For each Tweet T compute (opinion groups) do
5. Get adjective groups (AGi)
6. Get verb groups (VGi)
7. Count emoticons (Ne)
8. Calculate sentiment score of opinion groups
$$S(T) = \sum_{i=1}^{N(op)} (\text{opinion groups})$$
9. End for
10. Calculate overall sentiment score of tweet

$$S(T) = \frac{\left(1+\frac{Wc+Lr+En}{3}\right)}{N_{(op)}} * \sum_{i=1}^{OG(E)} S(AGi) + S(VGi) + Nei * S$$

11. Return Sentiment of Tweet,
12. if Abs (Score (T)) >Thr then
13. Return: T is subjective
14. Get (W, pol_score) from SWN;
15. for (i=1; i<=n; i++)
16. pos_score(W) ← pos_scorep(i) / n
17. neg_score(W) ← neg_scorep(i) / n
18. obj_score(W) ← obj_scorep(i) / n
19. End for
20. If pos_score(W) > neg_score(W)
21. max_pol_score(W) ← [pos_score(W)]
22. Else if neg_score(W) > pos_score(W)
23 max_pol_score(W) ← -[neg_score(W)]
24 Else
25. max_pol_score(W) ← [obj_score(W)p]
26. End if
27. if W preceded by NL then
28. max_pol_score (W) ← pol_score(W) * -1;
29. if W preceded by ERL then
30. max_ pol_score (W) ← pol_score(W) + getERL(enhancer_reducer_Word, score);
31. if W preceded by CSL then
32. max_pol_score (W) ← pol_score(W) + getCSL(context_shifter_Word, score);
33. Return: word sentiment score
34 End if
35. Else
36. Return: T is objective
37. End if
38. Return: T.S (Sentiment score).
39. End For

End begin

Overall Tweet Sentiment

Modifiers such as adverbs and adjectives are used to explain verbs and nouns respectively. Adjectives are grouped with noun and named as adjective group, while verbs and adverbs are combined to form verb group. Strength of the ith adjective group is given by the product of ith adjective score and noun score, similarly verb group strength is calculated by the product of verb score and adverb score. If adverb is not present in the opinion group then its default score will be 0.5.Theaverage of all the opinion intensifiers (capitalization, word emphasis, adjectives groups, verb groups, emoticons, exclamation mark) is calculated according to the formula given below:

$$S(T) = \frac{\left(1+\frac{Wc+Lr+En}{3}\right)}{N(op)} * \sum_{i=1}^{OG(E)} S(AGi) + S(VGi) + Nei * S(Ei) \qquad (3)$$

In equation 3, N (op) represents the number of opinion groups and emoticons in the tweet. “Wc”mentions the words fractions capitalized in the tweet. “Lr” shows the count of repeated letters. “En” shows the count of exclamation marks present in the tweet. W (AGi) denotes score of the ith adjective group, W(VGi) denotes the score of the ith verb group, W(Ei) denotes the score of the ith emoticon, Nei denotes the count of the ith emoticon,

Wc, Lr, En provide stress on the opinion to be conveyed and named as sentiment intensifiers. Tweets score is re-arranged to 1 and -1 if they exceed any one of the mentioned dimensions.

Weight of a Tweet

For the explanation of the proposed method, sentiment of the Tweet is calculated following the below methodology:
<tweet>=“@IRFAN… hate PMLN, its policies are HARAMFULLLL!!! I am totally in favour of PTI:(:(new party with new **EXPERIENCE”**

Tweet’s filtering andPointing Intensifiers and Opinion Groups

Filtering of tweets involve removing URL, Hash tags, targets and tweets special characters and is done by preprocessing module. As a result a file containing opinion barrier words is obtained.We compute scores of opinion intensifiers as follows:Total of eighteen words is in the tweet, out of which 4 words are capitalized. Named entity recognition module identify two of the capitalized characters as names, Therefore Wc=2/18=0.11

Length of the characters repeated multiple times Lr=4

Number of exclamation marks En=3

Tagging of the tweet is carried out with POS tagger and verb and adjective groups are extracted.

Adjective groups extracted are: (totally, favour), (policy, harmful), (new party), and (new experience).

Verb Groups identified in the tweet is (hate)

Emoticons are also included in the opinion group and its occurrence as well as types is counted. In the above tweet emoticon present is "**:(**" and its number of occurrence is 2.

Sentiment score of opinion groups

Adjective groups and verb groups are extracted, now we have to calculate their opinion score.

S (totally favour) =0.8*0.75 == 0.60

S (policy harmful) = -1*-.625= -0.625

S (new party) = 0.375*1 = 0.375

S (new experience) = 0.375*1 = 0.375

Semantic score of the verb identified is computed as below:

S (hate) = 0.5*-0.375 = -0.18

Calculating weight of a Tweet

Overall sentiment of the tweet can be calculated using the defined formula:

$$s(T) = .67 * \sum_{i=1}^{N(op)} (\text{opinion groups})$$

$$s(T) = .67 * ((.6) + (-.625) + (.375) + (.375) + 2 * (-.5)$$
$$= -.184$$

As the value contained is negative so the tweet can be placed in negative basket.

RESULT AND DISCUSSION

To analyze the mechanism and results of the adopted methodology we conduct an experiment. For evaluation we extract 521 political tweets from twitter public time line. The system executes the dataset and categorized tweets into 412 opinionative and 109 non-opinionative tweets where opinionative tweets are further classified into 221 positive and 109 negative tweets as well (Table 1). All the execution is carried out in accordance to the proposed methodology mentioned in methodology section. In Table 2 some tweets are placed along with their strength as well as with the assigned label.

Table 1.Categorization of opinionative and non-opinionative tweets

Dataset	User tweets	Opinionative tweets		Non-Opinionative tweets
		Positive tweets	**Negative tweets**	
Twitter Political tweets	**521**	**221**	**191**	**109**

Table 2: Tweets with sentiment score and label (positive, negative)

Political Tweets	**Sentiment Score**	**Label**
@beenasarwar1 Nov PPP: the MOOOST corrupt party…</3</3</3	**-1**	**Negative**
Performance of PPP was WORSSSST….!!!	**-0.475**	**Negative**
Every party just COLLLLECTS money @beenasarwar during the available TIME period… XD XDXD	**.758**	**Positive**
Political parties in Pakistan don't WORK for the WELFARE of the country….!!!	**.196**	**Neutral**
Give equal ATTEEEEEEEENTION to almost all the provinces….</3</3</3</3	**.643**	**Positive**
ttp://www.ndtv.com/article/world/I PMLN does't achieved its party MANIFESTO…..people DISAPPOINTED…:(:(:(:(	**-.811**	**Negative**
@ndtv: Stoooooopping NATO supply……:(:(:(… failed to prevent DRONE attttack	**-.627**	**Negative**
ttp://www.ndtv.com MQM is supporting terrorists in KARACHI….:'(:'(:'(	**-1**	**Negative**
PTI is going from BAD to WORSSSSSSST……X-(,X-(,X-(,X-(	**-1**	**Negative**

REFERENCES

[1] L. Colazzo, A. Molinari and N. Villa, "Collaboration vs. Participation: the Role of Virtual Communities in a Web 2.0 world", International Conference on Education Technology and Computer, pp.321-325, 2009.

[2] Castells, M., ."Communication, power and counter-power in the network society.", *International Journal of Communication*, **1**(1), pp. 238-266, 2007

[3] Wattal, S., D. Schuff, M. Mandviwalla, and C. Williams,."Web 2.0 and Politics: The 2008 U.S. Presidential Election and an E-Politics research agenda.", MIS Quarterly, **34**(4), pp.669-688, 2010.

[4] Honeycutt, C., and Herring, S., "Beyond Microblogging: Conversation and Collaboration via Twitter", Proceedings of the Forty-Second Hawaii International Conference on System Sciences (HICSS-42), 2009.

[5] Java, A., Song, X., Finin, T., and Tseng, B., "Why We Twitter: Understanding Microblogging Usage and Communities", Joint 9th WEBKDD and 1st SNA-KDD Workshop '07 , 2007.

[6] Bo Pang and Lilian Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In ACL-2004, 2004.

[7] lp.stanford.edu/courses/cs224n/2011/reports/patlai.pdf

[8] Kim, D., "Tweeting Politics: Examining the Motivations for Twitter Use and the Impact on Political

Participation", 61st Annual Conference of the International Communication Association, 2011.

[9] Pew Research Center, "Parsing Election Day Media: How the Midterms Message Varied by Platform", Pew, 2010, retrieved June 14, 2011, from http://pewresearch.org/pubs/1794/parsing-election-daymedia- messages-varied-by-platform

[10] Brendan O'Connor, RamnathBalasubramanyan, Bryan R. Routledge, and Noah A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In Proceedings of the International AAAI Conference on Weblogs and Social Media, 2010.

[11] Johan Bollen, Alberto Pepe, and Huina Mao.Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena, 2009. http://arxiv.org/abs/0911.1583.

[12] Philipp G. SandnerAndranik Tumasjan, Timm O. Sprenger and Isabell M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010.

[13] Jonathon Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classi_cation.In Proceedings of the ACL Student Research Workshop on - ACL, 2005.

[14] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In KDD, 2011.

[15] Kumar, Akshi, and Teeja Mary Sebastian. "Sentiment Analysis on Twitter."*IJCSI International Journal of Computer Science Issue*,**9.3**: 372-378(2012).

[16] POS Tagger: http://www.infogistics.com/textanalysis.html

[17] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives". In Proceedings of the Joint ACL/EACL Conference,2004, pp. 174–181

[18] WordNet: http://wordnet.princeton.edu/

[19] Muhammad Z. Asghar, M. Qasim, B. Ahmad, Sh. Ahmad, A. Khan, Imran A. Khan. Health miner: opinion extraction from user generated health reviews. International Journal of Academic Research Part A; 2013; 5(6),

[20] Muhammad Z. Asghar, A. Khan, Fazal M. Kundi, M. Qasim, F. Khan, R. Ullah, Irfan U. Nawaz. Medical opinion lexicon: an incremental model for mining health reviews. International Journal of Academic Research Part A; 2014; 6(1), 295-302.