Aspect Level Information Retrieval System for Micro Blogging Site

Manisha Punn[1], Nidhi Bhalla[2]

[1, 2]Swami Vivekanand Institute of Engineering and Technology, Banur, Punjab, India

Abstract*: In this paper we have calculated the polarity of people on two current affairs. For this, the tweets posted by twitter handler network friends and friends of friends on current affairs are extracted from twitter. Following this, two bags of words are created. One is positive bag of words and other is negative bag of words. The objective aspect based polarity evaluation is implemented. The polarity of tweets have illustrated and interpreted with application to understand current digital word in terms of its aspect, perspective by calculating polarity with parameter influencing thoughts, concepts, ideas and aspects of people expressing in digital world*

Keywords*:* Aspect, Information retrieval, Opinion mining, Perspective, Tweets

1. Introduction

Information retrieval deals with acquiring relevant information from a collection of information resources. After retrieving the information about web objects, the information is indexed and then ranked. Aspect[1] is a related object or member or part of a topic of interest on which the evaluation is made. Aspect level information retrieval system retrieves the information based on his/her aspect, perspective etc.

The present paper retrieves aspects of five celebrities on five different topics and using them proposes an algorithm for calculating the polarity of users' aspects. The polarity is calculated as either positive polarity or negative polarity.

2. Previous Work

In a research work carried by [2], a novel unsupervised model known as Cross-Perspective topic model was proposed for contrastive opinion mining. The opinions on any ad hoc query were determined based on learned parameters. This model quantified the difference in user perspectives. The research was carried out on two datasets, that is, statement records of U.S senators and world news report from three media in U.S, China and India. It is concluded that the proposed model showed effective results.

Another study performed by [3] proposed a feature wise opinion mining system which extracted features from user generated contents and then determined the intensity of those opinions. The positive and negative features were identified by extracting the associated modifiers and opinions. It was concluded that the recall value was lower than precision which means that certain correct feature-opinion pairs could not be recognized\ by the system correctly. Also, more number of features could not be extracted in negative list as most of the reviews written by users were on positive side.

3. Methodology of Proposed Work

Following steps were performed in order to carry out the proposed work as follows:

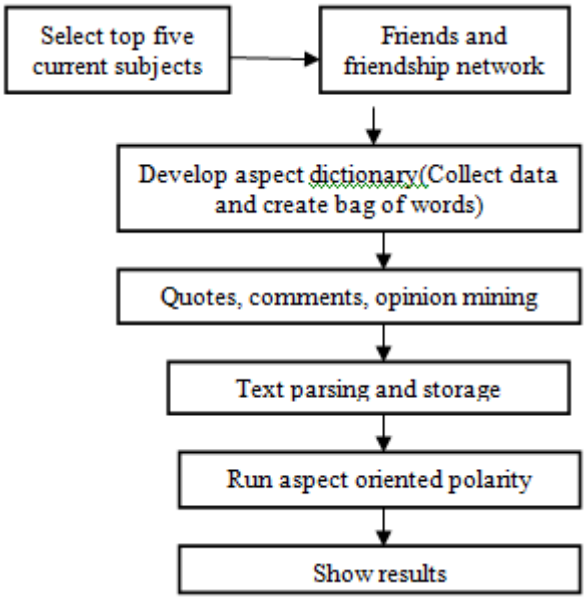

Figure 1: Methodology of proposed work

3.1. Select top 5 current subjects

The first step is to select the top five current subjects. The subjects were selected using Google Trends [4]. Google Trends show the total search volume of a particular search term across the world.

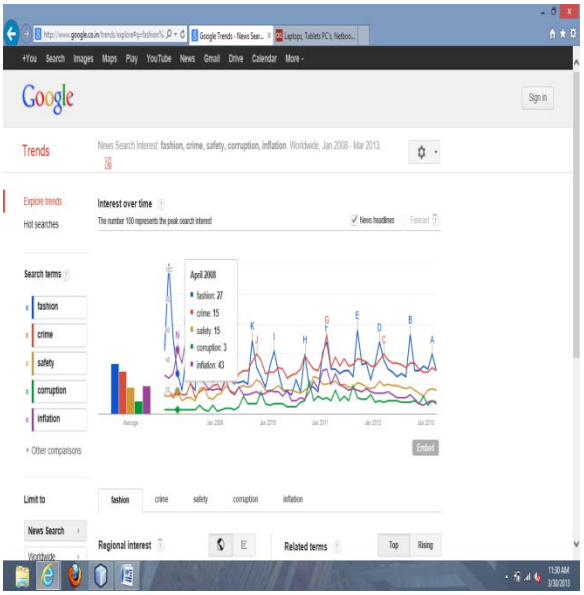

Figure 2: Selection of top five topics

3.2. Friends and Friendship Network

After selecting the top five current subjects, the friends and their friends are chosen whose tweets will be retrieved in the further steps.

3.3. Develop aspect dictionary (Collect data and create bag of aspect words)

This step deals with creating two separate aspect dictionaries. One contains all positive words and another contains all negative words.

3.4. Quotes, comments, opinion, extraction mining

In this phase, the tweets posted by friends and friendship network on current affairs are extracted from twitter. The current affairs included are:

a. Fashion
b. Crime
c. Safety
d. Corruption
e. Inflation

Table 1: Formulas to find positive or negative polarity

Concept	Mathematical Expression

Fashion	Total tweets related of Fashion concept-F
	Positive Dictionary=Pf
	Negative Dictionary=Nf
	x words € Pf (x-no of words in Pf)
	y words € Nf (y-no of words in Nf)
	Total words belong to tweet (T1)
	m positive words € Pf
	RATIO 1= x/m=>total words in Pf/ positive words
	RATIO 2=R1/F= (x/m)/F
	Rp=Ratio for every tweet in positive dictionary=((x/m)/F)*n
	∞
	Positive score'Ps'= sum of all ratios= ∑((x/m)/F)*n
	n=1
	Rn= Ratio for every tweet in negative dictionary= ((y/l)/F)*n
	∞
	Negative score'Ns'=sum of all ratios= ∑ ((y/l)/F)*n
	n=1
	If(Ps>Ns)-The person has PositivePolarity/views
	If (Ns>Ps)-The person has Negative polarity/views

Safety	Total tweets related to Safety-Sa
	Positive Dictionary=Psa
	Negative Dictionary=Nsa
	x words € Psa (x-no of words in Psa)
	y words € Nsa(y-no of words in Nsa)
	Total words belong to tweet (T1)
	o positive words € Psa
	RATIO 1= x/o => total words in Psa/positive words
	RATIO 2=R1/Sa= (x/o)/Sa
	Rp=Ratio for every tweet in positive dictionary=((x/o)/Sa)*n
	∞
	Positive score'Ps'= sum of all ratio= ∑((x/o)/Sa)*n
	n=1
	Rn= Ratio for every tweet in negative dictionary= ((y/r)/Sa)*n
	∞
	Negative score'Ns'= sum of all ratios= ∑ (y/r)/S)*n
	n=1
	If(Ps>Ns)-the person has Positive Polarity/views
	If(Ns>Ps)-The person has Negative polarity/views

Crime	Total tweets related to Crime-C
	Positive Dictionary=Pc
	Negative Dictionary=Nc
	x words € Pc (x-no of words in Pc)
	y words € Nc (y-no of words in Nc)
	Total words belong to tweet (T1)
	s positive words € Pc
	RATIO 1= x/s => total words in Pc/positive words
	RATIO 2=R1/C= (x/s)/C
	Rp=Ratio for every tweet in positive dictionary=((x/s)/C)
	∞
	Positive score'Ps'= sum of all ratio= ∑((x/s)/C)*n
	n=1
	Rn= Ratio for every tweet in negative dictionary= ((y/t)/C)*n
	∞
	Negative score'Ns'= sum of all ratios= ∑ ((y/t)/C)*n
	n=1
	If(Ps>Ns)-The person has Positive Polarity/views
	If(Ns>Ps)-The person has Negative polarity/views

Corruptio	Total tweets related to Fashion-CO
	Positive Dictionary=Pco
	Negative Dictionary=Nco
	x words € Pco (x-no of words in Pco)
	y words € Nco (y-no of words in Nco)
	Total words belong to tweet (T1)
	v positive words € Pco
	RATIO 1= x/v => total words in Pco/positive words
	RATIO 2=R1/co= (x/v)/CO
	Rp=Ratio for every tweet in positive dictionary=((x/v)/co)*n
	∞
	Positive score 'Ps'= sum of all ratio= ∑((x/v)/CO)*n
	n=1
	Rn= Ratio for every tweet in negative dictionary= ((y/w)/CO)*n
	∞
	Negative score 'Ns'= sum of all ratios= ∑ ((y/w)/CO)*n
	n=1
	If(Ps>Ns)-The person has Positive Polarity/views
	If(Ns>Ps)-The person has Negative polarity/views

Inflation	Total tweets related to Fashion-I
	Positive Dictionary=Pi
	Negative Dictionary=Ni
	x words € Pi (x-no of words in Pi)
	y words € Ni (y-no of words in Ni)
	Total words belong to tweet (T1)
	q positive words € Pi
	RATIO 1= x/q => total words in Pi/positive words
	RATIO 2=R1/I= (x/q)/Pi
	Rp=Ratio for every tweet in positive dictionary=((x/q)/I)*n
	∞
	Positive score 'Ps'= sum of all ratio= ∑((x/q)/I)*n
	n=1
	Rn= Ratio for every tweet in negative dictionary= ((y/k)/I)*n
	∞
	Negative score'Ns'=sum of all ratios= ∑ ((y/k)/I*)n
	n=1
	If(Ps>Ns)- The person has Positive Polarity/views
	If(Ns>Ps)-Negative polarity/views

3.5. Text parsing and storage

In this phase, pre processing is done and bad words are removed from the bag of words which was created in the previous step. After performing the pre processing, the data is stored in the local text database.

3.6. Run polarity evaluation algorithm

This phase deals with implementing the polarity evaluation algorithm. The input to the polarity evaluation algorithm is number and nature of positive and negative extracted, parsed, Tweets dataset and the output of this algorithm gives the polarity value in percentage.

4. Results

4.1 Fashion

Results show that the positive polarity of people on fashion is 11% and negative polarity is 89%.

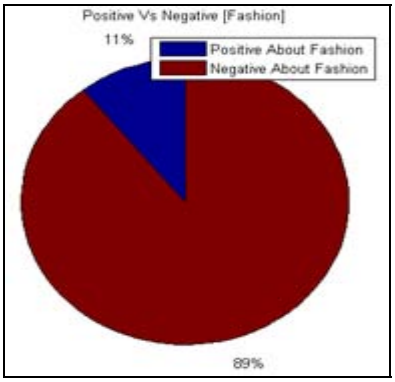

Figure 3: Polarity values of fashion

4.2 Safety

The results show that people's views on safety have positive polarity of 86% and negative polarity of 14%.

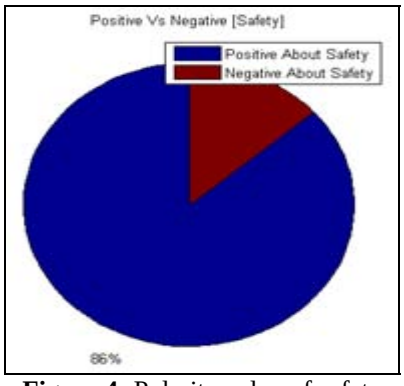

Figure 4: Polarity value of safety

4.3 Corruption

The results show that positive polarity of people on corruption is 9% and negative polarity is 91%.

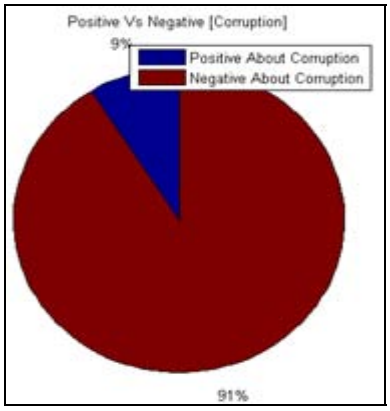

Figure 5: Polarity values of corruption

4.4 Crime

Results show that positive polarity of people on crime is 10% and negative polarity is 90%.

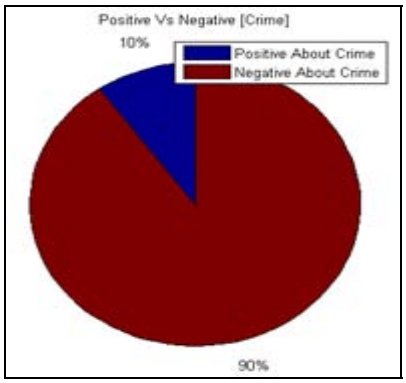

Figure 6: Polarity values of crime

4.5 Inflation

Results show that people's views on inflation have positive polarity of 6% and negative polarity of 94%.

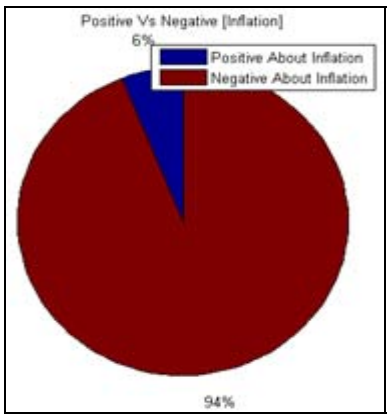

Figure 7: Polarity values of inflation

5. Conclusion

The objective of this research is to calculate the polarity of tweets posted by the celebrities. For calculating polarity tweets posted by friends and friendship network were retrieved following which bag of words were created. The polarity was calculated using an algorithm as mentioned in section. It is an objective evaluation which can be done without active participation of the people involved, and there

is no breach in privacy as only public dataset of tweets is processed, and it can run continuously to bring the statistical information , what percentage of people are thinking in what way-positive or negative. It is from this views we can access the content of thoughts and their flow of attributes, feelings, mood direction of the people towards the current affairs topics like fashion, crime, safety, corruption and inflation.

Future Scope

In our current research we have tried to capture the views, expressions and thoughts of people using digital tools like twitter and have formulated mathematical expressions for calculating their positive aspects and negative aspects of their expressions (tweets) in an non-invasive way, which produced fairly good realistic view of how people are thinking on current aspects of things happening around them, for future scope however we suggest a deeper analysis may be conducted to extend this work. The further analysis may consist of identifying trends between various aspects of people's opinions with identifying causal relationships between influencing parameters. This may be conducted using regression also.

References

[1] Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto, "Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining", 2007.
[2] Yi Fang, Luo Si, Naveen Somasundaram, Zhengtao Yu, "Mining Contrastive Opinions on Political Text Using Cross-Perspective Topic Model', 2012.
[3] Tanvir Ahmad, Mohammad NajmudDoja, "Ranking Sysem for Opinion Mining of Features from Review Documents", 2012.
[4] http://www.google.com/trends/

Authors Profile

Manisha Punn is pursuing M. Tech (Computer Science) form Swami Vivekanand Institute of Engineering and Technology, Banur, Punjab, India. She did her B. Tech (Information Technology) form Doaba Institute of Engineering and Technology, Kharar, Punjab, India.

Nidhi Bhalla is working as Assistant Professor in Swami Vivekanand Institute of Engineering and Technology, Banur, Punjab, India. She did her B. Tech (Information Technology) and M. Tech (Computer Science) form Lovely Professional University, Jalandhar, Punjab, India.