

Identification and Opinion Extraction through User Generated Content on Web Based Social Media

Dr. Deepak Arora, Kapil Verma

Professor, Dept. of computer science and Engineering (ASET) Amity University Uttar Pradesh Lucknow, India

Sr. Lecturer, Dept. of computer science and Engineering Babu Banarasi Das Northern India Institute of Technology (BBDNIIT), Uttar Pradesh, Lucknow, India

Abstract

Nowadays internet is becoming a platform where different user can post their ideas and opinions. The social networking sites and blogs offer a wide variety of such informative text which can be used to establish or determine a mindset for a particular product, person or individual. These blogs can be used as a vast source of information through which one can predict opinion as well as planning for different business strategies. Due to huge amount of information there is always need of specific tool or approach to mine useful text called opinion. Authors have proposed an approach of mining and classification for different real time datasets gathered from various sources of information, freely available on internet. Authors have tested the approach over these datasets and found suitable results. In this paper we propose a method that classifies a user-generated content on the basis of positive, negative, neutral, double negative, negative positive, triple negative. Authors have proposed rules for analyzing ideas and tested against dataset using Naive Bayes and Support Vector machine (SVM) model for accuracy and found best result 80.39 % for NB and 81.37 % for SVM.

Index Terms— Double Negative, Triple Negative, Negative Positive, 64 rules for polarity.

I. INTRODUCTION

Public sentiment is everything. With public sentiment, nothing can fail. Without it, nothing can succeed.

Abraham Lincoln

Opinion mining has now become very popular research field. Many research papers have already been published on various national and international journals and many researchers have made profuse and significant progress in opinion mining and still this research is in progress, there is always need of specific tool or approach to mine the useful text called opinion. Author proposed an approach which is useful for determining the orientation of user generated content on web based social media. Web based social media offer a wide variety of informative text, for example reviews about any product, views about an individual, liking and disliking of services etc. These information can be used for a variety of purposes in different area. Product reviews are helpful for customers. When individual go to shop any product they want reviews of other customers about the product. These reviews are helpful for determining mindset for particular product. Reviews would be more helpful for business purpose when machine correctly understand the natural language and sentiments of human in text. There are various sites (platform) where people can express their opinion and share ideas about any topic. There exist some opinion mining software and tools

in the market that gives opinion for the keyword searched but that are not much effective. Often it

analyze for positive, negative and neutral words in the sentence. In this paper authors have taken up those sentences that are constructed using positive, negative, neutral, double negative, negative positive and triple negative. Extract public opinion from the user generated content [9]. The basic work included in analysis of sentences is to classify in to positive, negative or neutral [2]. A different approach is used in determining the sentiment of sentence. Instead of classifying sentence only for positive, negative or neutral, we are using six parameters for classifying the sentiments of sentences: negative, positive, neutral, double negative, negative positive and triple negative. Some time the meaning of word is changed completely when using the two or three polarity words together in the same sentence. On the basis of these six parameters we can easily determine the sentiment of sentence. For example,

- 1- Afzal Guru didn't kill anybody.
- 2- Child actor/actress doesn't dislike not going to school.
- 3- I hate it whenever fake people say they hate fake people. It is like a double negative..... Isn't it??
- 4- "Didn't Have Nothing To Do With Drugs"
- 5- Breast Cancer, "I Won't Back Down!"
- 6- When I was a commissioner there I never did nothing nowhere near that.

7- The blanket didn't barely protect their shivering bodies.

The above seven sentences are double negative and triple negative sentences. If you look at first and second sentence "Afzal Guru didn't kill anybody", "Child actor/actress don't dislike not going to school". The second sentence actually means that the child actor/actress dislike going to school but there are three negative words which change the focus that whether the child actor/actress actually likes to go to school or not. In the first sentence two negative words used together so it basically become positive, just like calculation in mathematics. "Afzal guru didn't kill nobody" would mean that the man did kill someone. In the above other sentences three, four, five, six, and seven. These sentences also contain double or triple negative words.

In negative positive words a negative word comes just before positive word and changes its meaning from positive to negative. For example, "The quality of 'X' mobile phone Camera is not good."

This review sentence contain one positive word and its meaning changes from positive to negative because a negative word 'not' is written before it. List of positive words and negative words is given[6]. In this paper we particularly focus on six parameter and rules on the basis of that we determine the polarity of sentences. In section 2 related work and source of motivation for the ongoing research is discussed. Section 3 explains the process of mining reviews. Section 4 presents Method and rules that determine the polarity of sentence. In section 5 we discuss the findings of this research and result are discuss in section 6.

II. RELATED WORK

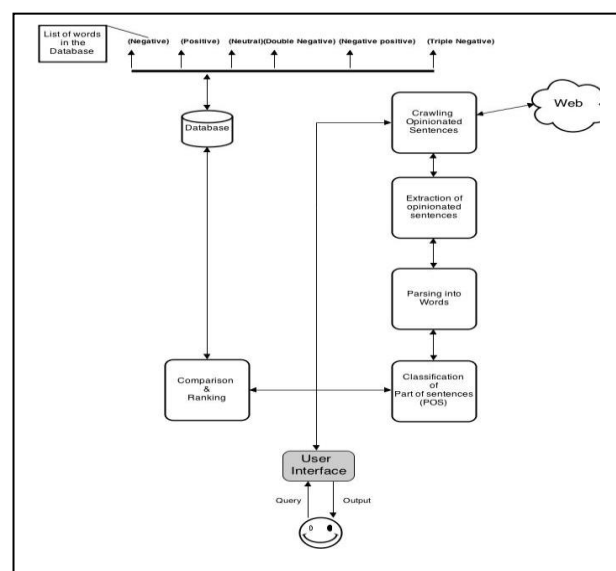
Social media provide abundance of information in the form of customer review. It is one of the current hot research topic for many researchers. Foundation and trends in information retrieval [3] discussed it in details. In there research, the author describe opinion oriented classification and opinion oriented Information retrieval. Classifying the opinion document as positive and negative [5]. Method to collect a corpus with positive and negative sentiments is given in [1]. In this author explain the method which allow to collect negative and positive sentiments such that no human effort is needed for classifying the document. Supervised learning model to determine if sentiment expressed on different topics in a conditional sentence [8] are positive, negative or neutral. Conditional sentence are the sentences that describe the hypothetical situation and there consequences. In my work sentence is analyze on the basis of six parameters and sixty four rule and exceptions governing to decide polarity of sentences.

Other similar work comes from feature base sentiment analysis [4] in this work the product feature such as design, quality etc of any product and opinion have been expressed on them. Some times opinion is positive on one feature and negative on other.

The subjective orientation have been studied [10] uses the adjective to find the sentence subjectivity. Hatzivassiloglou and wiebe [11] works on sentence subjectivity. They propose a method that search for adjective then reflect the negative or positive orientation. Comparative sentence is also discussed in the study [7]. Most of the user on internet wants to post an idea and opinion on the internet. Daily new forums are created to discuss on various topics. If any new product launched, people start discussing about it by sharing ideas and opinion in their own way on internet forum. Some express anger, flaws, some speak in favors of etc. Most of the time a biased decision has been taken due to limited accessibility of opinion in the form of reviews. So there is always a need of specific tool or approach to mine the useful text called opinion. For this reason it is necessary to find some approach that understand natural language and sentiment expressed by human in text. Many researcher is collectively doing research in this field for making automated system that understand exactly the same as human being. Many researchers actively working on it in many countries all over the world.

III. MODEL FOR OPINION MINING PROCESS

The following figure depicting the overall process of opinion mining-



In the above figure user put keyword in the user interface in response user interface process the query and start crawling the opinion based sentences from the web. It search opinionated sentences from forum, blogs, social networking sites and extracted

that opinionated sentences. Then it performs the parsing of sentences in to the words. Then perform classification of part of sentences (POS). Comparing it with the list of words for six parameters then we rank the word for each parameter in zero and one for the parameter present or absent in the sentence. A related work is done in opinion retrieval from blog [12]. Then by ranking method, sixty four rules and exceptions we find the orientation of the sentences and give result through the user interface.

IV. METHOD AND RULES FOR DETERMINING POLARITY OF SENTENCES.

In this paper authors are using six parameters for identifying the sentiment of sentences. These six parameter are:

Negative, Positive, Neutral, Double Negative, Negative Positive, & Triple negative and their codes are ‘a’, ‘b’, ‘c’, ‘d’, ‘e’, & ‘f’ respectively.

A. Negative Parameter

A parameter that expresses, containing, or consisting of negation. It refuses and denying the request and gives negative answer. e.g. no, not, never, scary etc. In this paper author uses code ‘a’ for negative parameter.

B. Positive Parameter

A parameter that shows acceptance, affirmation and positive response of a product, appraisal of service .e.g. good ,graceful ,happy etc. In this paper author uses code ‘b’ for positive parameter.

C. Neutral Parameter

A parameter that expresses no orientation.e.g. person, nation, country etc. In this paper author uses code ‘c’ for neutral parameter.

D. Double Negative Parameter

Double negative is used as a positive parameter. It gives the positive orientation. Basically the construction of double negative parameter is –

$$a + a = b. \quad (1)$$

Note: The above equation it is just like math when two negative parameter added it will give positive parameter.

E. Negative Positive Parameter

A negative positive parameter is used as a negative parameter. It gives the negative sense .The basic construction of negative positive parameter is –

$$a * b = a. \quad (2)$$

F. Triple Negative Parameter

A triple negative parameter is used as a negative parameter. When negative parameter is used three

times. First two negative will become positive by “Eq . 1” and when third negative is used with this positive it will become negative by “Eq. 2” .The basic construction of triple negative parameter is –

$$a * a * a = b \quad (3)$$

A. Determining The Polarity Of Sentences

The polarity of sentence is determined by the six parameter in the “TABLE I.”.Analysis of sentences is done by marking it 0 and 1.Mark 1 to those parameter if the sentence have word for that parameter. Mark 0 to those parameter if the sentence have no word for that parameter. Result is declared in the last column.

B. Rules For Determining The Polarity

There are total six parameter for the value 0 and 1.So the combination six parameter will be sixty four.

2 raise to the power 6 i.e. $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$.

Result is declared in the last column as negative (-ve), positive (+ve) and Neutral. Exceptions (Excp.) are also there in the result column.

So there will be total sixty four rule for determining the polarity are given in the TABLE I below.

TABLE I. 64 Rules For Polarity

SI No	Six Parameter						Result
	a	b	c	d	e	f	
Rule1	0	0	0	0	0	0	Neutral
Rule2	0	0	0	0	0	1	-ve
Rule3	0	0	0	0	1	0	-ve
Rule4	0	0	0	0	1	1	-ve
Rule5	0	0	0	1	0	0	+ve
Rule6	0	0	0	1	0	1	Excp.
Rule7	0	0	0	1	1	0	-ve
Rule8	0	0	0	1	1	1	Excp.
Rule9	0	0	1	0	0	0	Neutral
Rule10	0	0	1	0	0	1	-ve
Rule11	0	0	1	0	1	0	-ve
Rule12	0	0	1	0	1	1	+ve
Rule13	0	0	1	1	0	0	+ve
Rule14	0	0	1	1	0	1	Excp.
Rule15	0	0	1	1	1	0	Excp.
Rule16	0	0	1	1	1	1	Excp.
Rule17	0	1	0	0	0	0	+ve
Rule18	0	1	0	0	0	1	-ve
Rule19	0	1	0	0	1	0	-ve
Rule20	0	1	0	0	1	1	+ve
Rule21	0	1	0	1	0	0	+ve
Rule22	0	1	0	1	0	1	Excp.
Rule23	0	1	0	1	1	0	Excp.
Rule24	0	1	0	1	1	1	Excp.
Rule25	0	1	1	0	0	0	+ve
Rule26	0	1	1	0	0	1	Excp.
Rule27	0	1	1	0	1	0	Excp.

Rule28	0	1	1	0	1	1	Excp.
Rule29	0	1	1	1	0	0	+ve
Rule30	0	1	1	1	0	1	Excp.
Rule31	0	1	1	1	1	0	Excp.
Rule32	0	1	1	1	1	1	Excp.
Rule33	1	0	0	0	0	0	-ve
Rule34	1	0	0	0	0	1	+ve
Rule35	1	0	0	0	1	0	+ve
Rule 36	1	0	0	0	1	1	-ve
Rule37	1	0	0	1	0	0	-ve
Rule38	1	0	0	1	0	1	Excp.
Rule39	1	0	0	1	1	0	Excp.
Rule40	1	0	0	1	1	1	Excp.
Rule41	1	0	1	0	0	0	-ve
Rule42	1	0	1	0	0	1	Excp.
Rule43	1	0	1	0	1	0	Excp.
Rule44	1	0	1	0	1	1	-ve
Rule45	1	0	1	1	0	0	Excp.
Rule46	1	0	1	1	0	1	Excp.
Rule47	1	0	1	1	1	0	Excp.
Rule48	1	0	1	1	1	1	-ve
Rule49	1	1	0	0	0	0	-ve
Rule50	1	1	0	0	0	1	Excp.
Rule51	1	1	0	0	1	0	Excp.
Rule52	1	1	0	0	1	1	-ve
Rule53	1	1	0	1	0	0	Excp.
Rule54	1	1	0	1	0	1	Excp.
Rule55	1	1	0	1	1	0	Excp.
Rule56	1	1	0	1	1	1	-ve
Rule57	1	1	1	0	0	0	-ve
Rule58	1	1	1	0	0	1	Excp.
Rule59	1	1	1	0	1	0	Excp.
Rule60	1	1	1	0	1	1	-ve
Rule61	1	1	1	1	0	0	Excp.
Rule62	1	1	1	1	0	1	Excp.
Rule63	1	1	1	1	1	0	Excp.
Rule64	1	1	1	1	1	1	-ve

The above sixty four rules are sufficient for determining the polarity of sentence. There are exceptions in the sixty four rules because some of rules are not sufficient for directly determining the polarity of sentence. So thirty two exceptions are included in sixty four rules.

Exception rule are made based on seven coordinating conjunction (CC) and some subordinating conjunction (SC) which are given below.

Coordinating conjunction are AND, Or, But, Nor, So, For and Yet. Subordinating conjunction are Even Though, Even If, Once, Provided, and While.

Conjunctions are used to join part of sentences (POS). There are seven coordinating conjunction used to join part of sentences (POS). Below, discussing seven CC and some SC.

AND: It is very common conjunction used to join POS. It is observed that one POS is the result of another POS. So the polarity word appears after AND will be the result.

OR: It is also very common conjunction use to show possibility, correction, and polarity condition. For example, "UPA must decrease inflation or they would not be in 2014".

BUT: It is used to connect to two clauses with contrast in there meaning. Scan the part of sentence after but conjunction .The polarity word used after BUT will be the result.

NOR: It s generally used with neither nor. Mostly it is used with negative.It takes the sentiment of sentences towards negative.

SO: It is used to show the reflection of two individual clauses. It gives the over all sentiment of sentence. So the sentiment word used after SO will decide the result.

FOR: It is a conjunction use to give reason for the POS use before for. So result will be decided by sentiment word in POS before FOR.

YET: It has several uses. We can use it as still, adding, even etc. For Example, "public complaining loudly about increasing rape cases, yet government not taking any action".

Even though: It is subordinating conjunction. Its projection is based on first part. For Example "X is my good friend, even though X is smoker".

Even IF: It is often use to give condition in sentences. For Example, "Credibility of Nokia will not decrease even if one of its products fails".

Once: It has ambiguity. Some times it use in respect of time and some time it is use as situation. For example , " I'll go and watch movie, once I finish my dissertation work". First half of the sentences is depend on the second half.

Provided: It gives the condition. First half is depend on the second half. For Example, "BSNL is good company, provided it improves services".

While: It is also a condition. It is observed orientation of sentences is based on sentiment word after while. For example, "I'll stay here while raining".

V. RESULT AND DISCUSSIONS

A.DATA SETS

In our research, author have presented the six parameters for analysis of sentences, model for opinion mining process and method and rule for determining polarity of sentences. We use dataset from the different sources freely available on internet. Our main consideration in collecting dataset is to include actual data from different social networking sites, blogs, and other review related sites. We have also taken sentences that contain six parameters and also contain conjunction in the sentences. We then construe many sentences from different sources. We

also construe the conjunction in the sentences .We have noted pattern use in sentences. In our observation, we noted pattern in which people use tag line of product instead name of product to comment. Many of the sentences have no orientation i.e. positive, negative. We put that sentence in the neutral category. But our main consideration in this research is to analyze the sentences on the basis of six parameter and conjunction. We have construed our analysis of sentence nearly around 102 selected sentences. In section IV we declare the six parameters. Table I contain the sixty four rule. It includes all the possible combination of six parameters. We used the kappa score for computing inter rated agreement. We achieved two score 0.6663 for naïve bayes classifier and 0.6891 for support vector machine(SVM).This shows SVM has strong agreement than naïve bayes .

B.EXPERIMENTAL RESULT

Now we show the experimental result of user generated content. We compare the result of naïve bayes and SVM implementation, which shows the effective result. We use 10 fold cross validation for the best result.

Three Class classifications:

We have taken result for two different classifiers. One is naïve bayes and other is Support vector machine (SVM).We use negative, positive, and neutral as three class classification. We have shown the result in Table III for two classifiers. We have computed TP Rate, FP Rate, Precision, Recall, F-Measure and ROC Area for three classes.

For this we have taken 102 sentences to perform experiment. Although accuracy for the model naïve bayes is 80.39 % and SVM 81.37 %.which yielded accuracy for naïve bayes is lower. Table II shows the confusion matrix for Naïve bayes and SVM respectively. This shows that, from out of 102 sample sentences 42 are negative, 48 are positive and 12 are negative.

TABLE II. Confusion Matrix

Confusion Matrix NB			
Classes	Negative	Positive	Neutral
Negative	29	12	1
Positive	5	43	0
Neutral	1	1	10
Confusion Matrix SVM			
Classes	Negative	Positive	Neutral
Negative	37	4	1
Positive	11	36	1
Neutral	1	1	10

In Table II confusion matrix, for NB of 42 actual negative, this model predicted 29 were negative, 12 were positive , and 1 were neutral. For SVM of 42 actual negative, this model predicted 37 were negative, 4 were positive, and 1 were neutral. Similarly we can see for other class positive and neutral. These conclude that confusion matrix for SVM is better because accuracy for SVM is 81.37 %.

TABLE III. Accuracy by class using NB & SVM

Accuracy by class using Naïve Bayes							
	TP Rate	FP Rate	Prec .	Rec .	F	ROC Area	Class
1	0.69	0.1	0.829	0.69	0.753	0.908	Negative
2	0.896	0.241	0.768	0.896	0.827	0.907	Positive
3	0.833	0.011	0.909	0.833	0.87	0.961	Neutral
4	0.804	0.156	0.809	0.804	0.802	0.914	Wt.Avg.
Accuracy by class using SVM							
	TP Rate	FP Rate	Prec .	Rec .	F	ROC Area	Class
1	0.881	0.2	0.755	0.881	0.813	0.876	Negative
2	0.75	0.093	0.878	0.75	0.809	0.866	Positive
3	0.833	0.022	0.833	0.833	0.833	0.943	Neutral
4	0.814	0.129	0.822	0.814	0.814	0.879	Wt.Avg.

Fig. 2. Shows the threshold curve for negative ,positive, and neutral .In Fig 2 a,b,c represents negative, positive , and neutral respectively for NB and d,e,f represents negative, positive , and neutral respectively for SVM .The X-Axis represents the FP-Rate and Y-Axis represent TP-Rate .Table III shows the accuracy for different accuracy measure's parameter .When we compare it with each other we found weighted average TP Rate for NB(Naïve Bayes) were 0.804 and for SVM were 0.814.Similarly FP Rate for NB were 0.156 and for SVM were 0.129.Precision and Recall for NB were 0.809 and 0.804 respectively and for SVM were 0.822 and 0.814 respectively. F measure for NB were 0.802 and for SVM were 0.814.It is clearly seen that SVM yielded

good accuracy. So we use SVM that gives best result using 10 foldcross validation.

So we finally found SVM yielded better result.

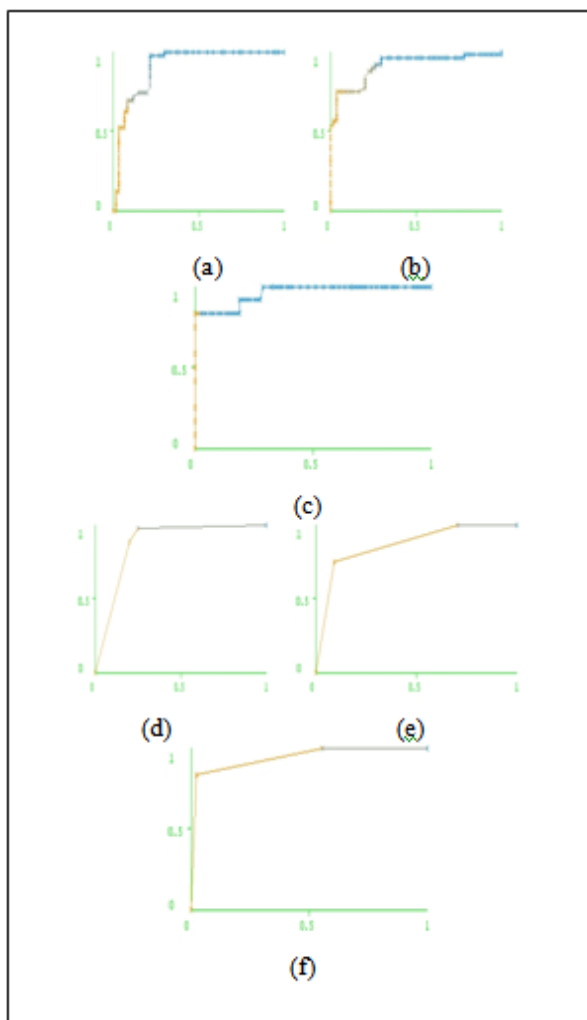


Fig. 2. Threshold curve For Negative Positive and Neutral. a,b,c represents negative ,positive and Neutral respectively for NB. andd,e,f represents negative ,positive, and neutral respectively for SVM.

VI. CONCLUSION AND FUTURE WORK

To perform Identification and Opinion Extraction through User generated Content on Web Based Social Media accurately, author propose and implement approach with the help of datasets. This paper studied sixty four rules for polarity and conjunction used in sentences. It is unlikely that there is common solution for all problems. Authors have taken seven coordinating conjunctions and some subordinating conjunctions. Our work was carried out by computational and linguistic pattern. In linguistic study, authors focused on sentence pattern after using conjunction, which have been explained through

useful result in section v. In computational study, authors have done classification through NB model and SVM model to predict accuracy of whether opinions on topics are positive, negative or neutral. Experimental found the best result for the datasets.

In Future work, author will improve the classification accuracy and study more in linguistic for identifying orientation of topic. Although there are many different conjunctions uses in the sentences, author will use all these conjunction in the sentences for determining the polarity of sentences.

VII.ACKNOWLEDGMENT

The authors are very thankful to their respected Mr. AseemChauhan, Chairman, Amity University, Lucknow, Maj. Gen. K.K. Ohri, AVSM (Retd.), Pro Vice Chancellor, Amity University, Lucknow, India, for providing excellent computation facilities in the University campus. Authors also pay their regards to Prof. S.T.H. Abidi, Director and Brig. U.K. Chopra, Deputy Director, Amity School of Engineering, Amity University, Lucknow for giving their moral support and help to carry out this research work.

REFERENCES

- [1] Alexander Pak and Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," In the proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC', may 2010), pp.347– 354.
- [2] Pang, B., Lee, L., and Vaithyanathan, S., "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2002, pp. 79-86.
- [3] B.Pang and L.Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information retrieval, vol.2, nos.1-2, 2008, pp.1-135.
- [4] Bing Liu, "Sentiment Analysis and Subjectivity, in NLP Handbook, 2nd Edition, N. Indurkha and F. J. Damerau, USA- Chapman and Hall/CRC, 2010, pp.
- [5] Bing Liu, "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012, pp. 30-43. R.Nicole, "Title of paper with only first word capitalized," J.NameStand. Abbrev., in press.
- [6] Hu and Liu, "A list of positive and negative opinion words or sentiment words for English," in KDD-2004
- [7] N jindal and B Liu. "Identifying Comparative Sentences in Text Documents. In Proc. Of the 29th SIGIR. 2006
- [8] Ramanathan Narayanan, Bing Liu, Alok Choudhary, "Sentiment analysis of

- conditional sentences*," in the proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1,pp. 180-189.
- [9] Ting-Chun Peng, Chia-Chun Shih., "Using Chinese part-of-speech patterns for sentiment phrase identification and opinion extraction in user generated reviews," in Proceedings of the 5th International Conference on Digital Information Management (ICDIM), 2010, pp. 120 - 127 .
- [10] V Hatzivassiloglou and K McKeown. Predicting the semantic orientation of adjectives. In *Proc. Of ACL. 1997*.
- [11] V Hatzivassiloglou and J Wiebe. *Effects of adjective orientation and gradability on sentence subjectivity. COLING'00. 2000.*
- [12] Wei Zhang, Clement Yu, Weiyi Meng, "Opinion retrieval from blogs," in the proceedings of the sixteenth ACM conference on Conference on information and knowledge management pp. 831-840