International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 5 | ISSN : 2456-3307

Identification of Opinion Words and Polarity of Reviews in Tweets using Aspect Based Opinion Mining

M. Vadivukarassi*, N. Puviarasan, P. Aruna

Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India

ABSTRACT

Opinion words extraction about products reviews from twitter is becoming an interesting area of research. This research motivated to develop an automatic opinion mining application for users. In this paper, a novel model for detecting noun phrases opinion words of user about the product reviews from tweets were proposed and aspect-based opinion mining model was used to get the polarity of opinion words as positive and negative reviews from Twitter dataset. The real-time datasets are analysed based on the experimental results.

Keywords: Aspect, Opinion mining, Polarity, Twitter

I. INTRODUCTION

Opinion and sentimental analysis technique is an efficient means of determining public opinions. Usually to collect customer comments, online or paper based surveys are used. Now, people tend to comment on their opinion in the social networking websites such as facebook or tweet profile. Therefore, the paper-based approach is not an efficient approach to review the customer comments. Therefore, Facebook, Twitter and all other social media sites are full of people's opinions about products/services they use, comment about popular personalities and much more. In this paper, large data sets from Twitter are analysed to determine the popularity of a given product from tweets. Tweets are a dependable source of information because people tweet about everything they do at present including buying new products and reviewing them. The data are collected using the Twitter public API, which allows developers to extract tweets from twitter programmatically. The unnecessary information is removed from the collected data [1].

The rest of the paper is organized as follows. The literature work regarding the twitter data analysis which used for determining the polarity of tweets has been discussed in "Related work". The proposed work of the aspect based opinion words identification has been implemented in "Proposed system". The results of the work have been visualized graphically and suggestion made along with supporting literature has been obtained in "Result and Discussion". Finally, "Conclusion" draws the conclusion.

II. RELATED WORKS

Jantima Polpinij et al. described towards the problem of analyzing the feedback in multiple languages and proposed the classification of multilingual sentiments analysis. The processing of framework consist two steps where first step deals with classification of reviews into two language classes and second step focuses on classifying textual dataset into positive and negative sentiments. The first step is implemented with the help of lingual separation of review by employing character analysis. The next step uses Latent Semantic Indexing method to group the similar words from group of documents based on word concept [2]. Samir E.Abdelrahman, Ebtsam Abdelhakam Sayed, Reem Bahga contributed about the integration of two SentiWordNet based on sentence level polarity classification. The main step is employed to find the score of foremost word sense using word sense disambiguation algorithm and to collect list of non-zero

values for prior sentiment words of subjectivity bound lexicons. The experiment had conducted based on different lexical resource that had merged with feature selection to train the classifier. The system fails to use best feature selection algorithm that results in improper classification of polarity of lexicon [3].

Akshil Kumar, Teeja Mary Sebastian proposed to mine the sentiment analysis of tweets from Twitter and developed a hybrid approach using both corpus based and dictionary based methods to determine the semantic direction of the opinion words in tweets [4]. Su Su Htay and Khin Thidar Lynn used the data of movie review, customer feedback review, and product review. The numerous statistical feature selection methods are used and directly machine learning techniques are applied. These experiments show that machine learning techniques only are not well performing on sentiment classification and show that the presence or absence of word seems to be more indicative of the content rather than the frequency for a word [5]. Zhang and Liu analysed to identify such opinionated noun features and this method to deal with the problem for finding product features, which are nouns, or noun phrases are proposed [6].

III. PROPOSED ASPECT BASED MINING

The aspect-based opinion mining is also referred as the feature-based opinion mining. The opinion goal has been decomposed into entity and its aspects. The aspects are used for representing the entity itself in the result and covers both entities and aspects. The target of opinion mining is to extract customer feedback data such as opinions on products and present information in the most effective way that serves the chosen objectives. Customers express their opinion words in review sentences with single word or phrase. Let us use an example of the review in the tweets: “Battery life is short.” In this sentence, the aspect (feature) is “battery life” and opinion word is “short.” Therefore, the aspects and opinion words need to be identified from the tweets. Figure 1 shows the overall process for generating the results of aspect-based opinion mining.

The system input is the real time user data about the products collected from the Twitter data. The preprocessing of the tweet is performed first for removing the needless symbols, then tweet tokenization is used to split the group of tweets into single tweet, POS tagging to parse the sentence and then identify product aspects and opinion words. The extracted opinion words are used to find out the opinion direction, which is positive or negative. Finally, the opinions for each product feature based on their orientations are summarised.

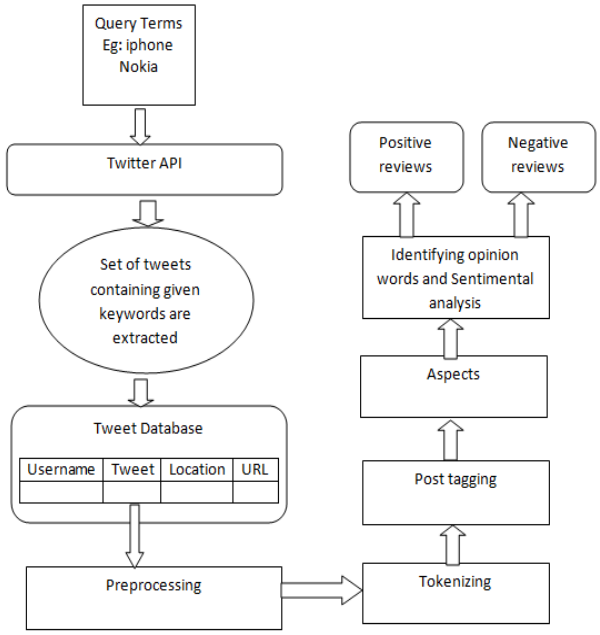


Figure 1. Overall Block Diagram of the proposed system

Steps of the Proposed System

Step 1: Extract the tweets from the database using certain keywords.
Step 2: Preprocessing steps such as stop word removal, white space removal, special symbol removal of the extracted reviews are done.
Step 3: Part of Speech tagging of the preprocessed tweets are performed.
Step 4: Opinion word list are created from the POS tagged reviews.
Step 5: Compute the score of word through SentiWordNet dictionary, where the score of all aspects that have same part of speech as of the word are computed and averaged to give final score of the word.
Step 6: The score of the aspect based review are calculate by using step 5.
Step 7: If calculated total score is > 0 then polarity of aspect based review is positive.
Else if calculated total score is <0 then polarity of aspect based review is negative.

A. Tweet Extraction

The tweets are used as a data set, which is used to extract tweets in a large scale from Twitter using the Twitter public API. This tweet extraction has been implemented in Python script. Twitter grants a large set of filtering parameters so that a well-defined set of tweets can be acquired using Twitter API.

Once the query has been created it can be ran by the API and all relevant twitter data information will be provided as output in the browser and this data was directly inserted into a database. Each tweet contains several types of information like user name, tweet id, text, location, url etc. Initially the twitter API allowed tweet locations in the form of latitude and longitude to be available with every tweet where the user has made his/her location public. The tweets are collected based on the keywords are extracted and saved in database.

B. Tweet Preprocessing

The data obtained from the API obviously contains many non-relevant information. Very basic and simple cleanup was performed using Python script. The random characters and other useless information in a tweet were filtered out before further analysis. Natural Language Processing techniques were used to filter out these useless data. In order to provide only important data, in general a clean tweet should not contain URLs, hashtags (i.e.#Rain) or mentions (i.e. @kamal).Then tabs and line breaks should be replaced with a blank and quotation marks with apexes. After, all the punctuation is removed in this step. \

C. Tweet Tokenization

Tokenization is the process of breaking a stream of text into meaningful words (stems), phrases or symbols. The tokens can be used further for parsing (syntactic analysis) or text mining. Tokenization is considered easy relative to other tasks in text mining. The first step is to segment text into words in majority of text processing applications. In English languages, word tokens are delimited by a blank space. Thus, for such languages, which are called segmented languages token boundary identification is a somewhat trivial task since the majority of tokens are bound by explicit separators like spaces and punctuation.

A simple token which replaces white spaces with word boundaries and cuts off leading and trailing quotation marks, parentheses and punctuation produces an acceptable performance. The next step is to handle abbreviations. In English languages, even though a period is directly attached to the previous word, it is usually a separate token which signals the termination of the sentence.

D. Part of Speech Tagging

The main goal of this paper is to find out product features and opinion words and also to find polarity of opinion word as positive and negative. In general, opinion words are adjectives and product features are nouns. Consider following example

"iphone 8 is excellent phone"

In above sentence, phone (product feature) is noun and excellent (opinion word) is adjective. In part-of-speech (POS tagging), each word in review is tagged with its part of speech (such as noun, adjective, adverb, verb etc). After POS tagging now it is possible to retrieve nouns as product features and adjectives as opinion words.

Figure 2 shows how above sentence will be tagged using POS Tagger. In tagged sentence, excellent is tagged with tag JJ which indicates 'excellent' is an adjective where a 'phone' is tagged as NN which indicates noun. Table 1 indicates the part of speech tagger

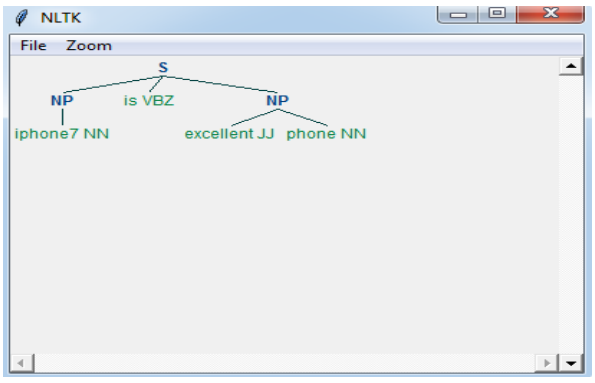


Figure 2. POS Tag parser structure

Table 1. Parts of Speech Tagger

Tag	Description
IN	Preposition or subordinating
JJS	Adjective, Comparative
JJR	Adjective, Superlative
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative

E. Aspect Extraction

The key characteristic of aspect extraction is that an opinion always has a target. The target is often word to be extracted from a sentence. Each opinion expression and its target from a sentence are recognized. Some opinion expressions can take part in two tasks, i.e., signifing a positive or negative polarity and implying an (implicit) aspect (target). For example, in "this mobile is cheap," "cheap" is a sentiment word and indicates the aspect price. Here, the approach called aspect extraction is used based on frequent nouns and noun phrases. Since existing research on aspect extraction is mainly carried out in online tweets and these tweets context to describe these techniques. In this paper, finding frequent nouns and noun phrases find explicit aspect expressions that are nouns and noun phrases from a large number of reviews in a given domain.

F. Identifying opinion words and Sentimental analysis

The opinion words are generally adjectives words. If a sentence contains one or more product features and one or more opinion words, then the sentence are called an opinion sentence. In opinion polarity identification word, semantic orientations of each opinion words are identified. Semantic orientation means identifying whether opinion word is expressing positive or negative opinion. Opinion polarity identification predicts the orientation of an opinion sentence. Consider following sentence-

"It is not an excellent phone"

The above sentence contains opinion word 'excellent' which expresses positive opinion. But sentence expresses negative opinion because of negation word 'not'. So by finding opinion word polarity identification it is necessary to find polarity of opinion sentence. A list of negation words such as 'no', 'not', 'but' etc. can be prepared and negation rules can be formed for finding the opinion sentence polarity identification. To determine the final orientation of the sentiment on each aspect in the sentence, the resulting sentiment scores are calculated using the equation (1). Let the tweet be s, which contains a set of aspects $\{as_1,\ldots, as_m\}$ and a set of sentiment words or phrases $\{sw_1, \ldots, sw_n\}$ with their sentiment scores. The sentiment orientation for each aspect as_i in s is determined by the following aggregation function:

$$Score(as_i, s) = \sum_{ow_j \in s} \frac{sw_j.so}{dist(sw_j, as_i)} \quad (1)$$

where sw_j is an sentiment word/phrase in s, $dist(sw_j, as_i)$ is the distance between aspect as_i and sentiment word sw_j in $s.sw_j$. so is the sentiment score of sw_i. The multiplicative inverse is used to give lower weights to sentiment words that are far away from aspect as_i. If the final score is positive, then the opinion on aspect as_i in s is positive. If the final score is negative, then the sentiment on the aspect is negative. Table 2 shows the tweet, which identifies the opinion words and its score value respectively.

Table 2. Identifying opinion words and its score value

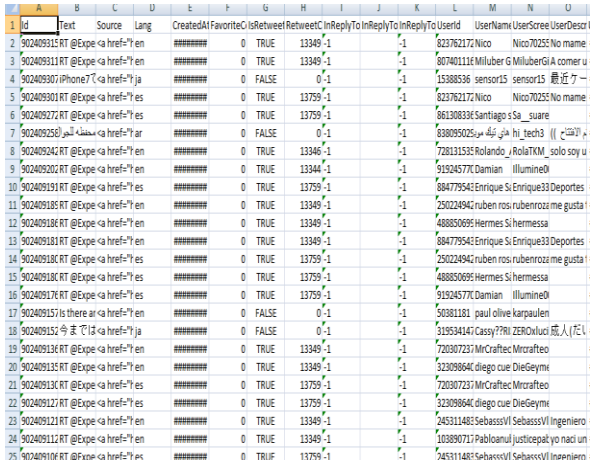
Word	POS tag	Score
Iphone	NN	0
Is	VBZ	0
Excellent	JJ	+2
Phone	NN	0
Overall Score		+2

IV. RESULTS AND DISCUSSION

A. Dataset Collection

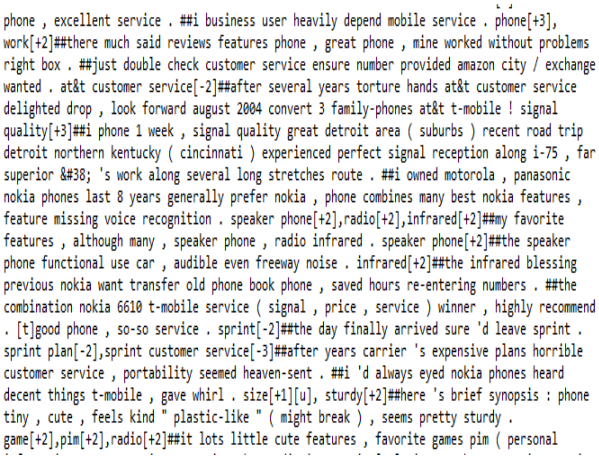
In this paper, the input to the opinion mining process is the set of real time tweets. Tweets are queried several times between July and August of 2017 to obtain information mentioning the mobile products. Twitter offers two APIs to retrieve data from tweets: REST and Streaming. In this paper, the REST API was used in Python script. The collected reviews are stored in a database, which is used for the Opinion Mining process. The only electronic product trending on twitter was the iPhone 8. So the product iPhone 8 are used for analyzing the tweets.

The tweets which contained the term ‘iPhone 8’ was collected from Twitter Streaming API and decided to determine which feature of the iPhone 8 was most or least popular the query was enhanced using a few keywords to obtain feature specific tweets. An example would be ‘iPhone 8 battery’. This query parameter will cause the API to return only tweets, which contain both iPhone 8, and battery terms together which results in tweets about the battery performance of the iPhone 8. Other keywords used were “camera”, “iOS”, “iTunes”, “screen”, “sound”, and “touch”. For each tweet, the user name, tweet text, location are extracted. Figure 3 represents the collection of tweets in a database. Figure 4 presents the text extraction from the database.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Id	Text	Source	Lang	CreatedAt	FavoriteC	IsRetweet	RetweetC	InReplyTo	InReplyTo	InReplyTo	UserId	UserNam	UserScree	UserDescr
2	902409315	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	823762172	Nico	Nico70255	No mame
3	902409311	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	807401116	Miluber G	MiluberGi	A comer u
4	902409307	iPhone7で	<a href=""	ja	########	0	FALSE	0	-1		-1	15388536	sensor15	sensor15	最近アー
5	902409301	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	823762172	Nico	Nico70255	No mame
6	902409272	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	861308336	Santiago s	Sa__suare	
7	902409256	لتحر محطة#	<a href=""	ar	########	0	FALSE	0	-1		-1	838095025	هاي تيك سر	hi_tech3	((الفتح
8	902409242	RT @Expe	<a href=""	en	########	0	TRUE	13346	-1		-1	728131535	Rolando ♪	RolaTKM_	solo soy u
9	902409202	RT @Expe	<a href=""	en	########	0	TRUE	13344	-1		-1	91924577	Damian	IIlumineO	
10	902409191	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	884779542	Enrique Si	Enrique33	Deportes
11	902409185	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	250224942	ruben ros	rubenroz	me gusta
12	902409186	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	488850695	Hermes Si	hermessa	
13	902409181	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	884779542	Enrique Si	Enrique33	Deportes
14	902409180	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	250224942	ruben ros	rubenroz	me gusta
15	902409180	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	488850695	Hermes Si	hermessa	
16	902409176	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	91924577	Damian	IIlumineO	
17	902409157	Is there ar	<a href=""	en	########	0	FALSE	0	-1		-1	50381181	paul olive	karpaulen	
18	902409152	今までは	<a href=""	ja	########	0	FALSE	0	-1		-1	319534147	Cassy??RI	ZERO×luci	成人(だい
19	902409136	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	720307237	MrCraftec	Mrcrafteo	
20	902409135	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	32309864	diego cue	DieGeym	
21	902409134	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	720307237	MrCraftec	Mrcrafteo	
22	902409127	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	32309864	diego cue	DieGeym	
23	902409121	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	245311483	Sebasssv	Vl Sebasss	Vl Ingeniero
24	902409112	RT @Expe	<a href=""	en	########	0	TRUE	13349	-1		-1	103890717	Pabloanul	justicepat	yo nacl un
25	902409106	RT @Expe	<a href=""	es	########	0	TRUE	13759	-1		-1	245311483	Sebasssv	Vl Sebasss	Vl Ingeniero

Figure 3. Collection of tweets in database

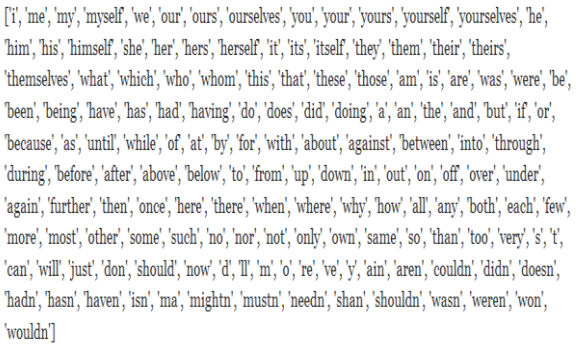


phone , excellent service . ##i business user heavily depend mobile service . phone[+3], work[+2]##there much said reviews features phone , great phone , mine worked without problems right box . ##just double check customer service ensure number provided amazon city / exchange wanted . at&t customer service[-2]##after several years torture hands at&t customer service delighted drop , look forward august 2004 convert 3 family-phones at&t t-mobile ! signal quality[+3]##i phone 1 week , signal quality great detroit area (suburbs) recent road trip detroit northern kentucky (cincinnati) experienced perfect signal reception along i-75 , far superior & 's work along several long stretches route . ##i owned motorola , panasonic nokia phones last 8 years generally prefer nokia , phone combines many best nokia features , feature missing voice recognition . speaker phone[+2],radio[+2],infrared[+2]##my favorite features , although many , speaker phone , radio infrared . speaker phone[+2]##the speaker phone functional use car , audible even freeway noise . infrared[+2]##the infrared blessing previous nokia want transfer old phone book phone , saved hours re-entering numbers . ##the combination nokia 6610 t-mobile service (signal , price , service) winner , highly recommend . [t]good phone , so-so service . sprint[-2]##the day finally arrived sure 'd leave sprint . sprint plan[-2],sprint customer service[-3]##after years carrier 's expensive plans horrible customer service , portability seemed heaven-sent . ##i 'd always eyed nokia phones heard decent things t-mobile , gave whirl . size[+1][u], sturdy[+2]##here 's brief synopsis : phone tiny , cute , feels kind " plastic-like " (might break) , seems pretty sturdy . game[+2],pim[+2],radio[+2]##it lots little cute features , favorite games pim (personal

Figure 4. Text extraction from database

B. Preprocessing

After data collection, the tweets are extracted separately from the text file. In this paper, the accuracy of data is improved by preprocessing method and avoids unnecessary data processing in each phase. It includes remove unnecessary words and non alphabetical characters. Here, NLTK techniques are used for preprocessing the tweets as shown in Figure 5.



['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', 'her', 'hers', 'herself', 'it', 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', 'should', 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', 'couldn', 'didn', 'doesn', 'hadn', 'hasn', 'haven', 'isn', 'ma', 'mightn', 'mustn', 'needn', 'shan', 'shouldn', 'wasn', 'weren', 'won', 'wouldn']

Figure 5. Some list of stop words

C. Tokenizing the tweet

This method will read reviews from text file after preprocessing the tweets. The text file consists of paragraphs. Therefore, the entire file will break into sentence. This sentence can individually used for mining as shown in Figure 6.

```
8   signal quality[+3]##i have had the phone for 1 week , the signal quality has been great in
the detroit area ( suburbs ) and in my recent road trip between detroit and northern kentucky (
cincinnati ) i experienced perfect signal and reception along i-75 , far superior to at &#38; t
's which does not work along several long stretches on that same route .
9   ##i have owned motorola , panasonic and nokia phones over the last 8 years and generally
prefer nokia , this phone combines many of the best nokia features , the only feature missing
for me is the voice recognition .
10  speaker phone[+2],radio[+2],infrared[+2]##my favorite features , although there are many ,
are the speaker phone , the radio and the infrared .
11  speaker phone[+2]##the speaker phone is very functional and i use it in the car , very
audible even with freeway noise .
12  infrared[+2]##the infrared is a blessing if you have a previous nokia and want to transfer
your old phone book to this phone , saved me hours of re-entering my numbers .
13  ##the combination of the nokia 6610 and t-mobile service ( signal , price , service ) is a
winner , i highly recommend it .
14  [t]good phone , so-so service .
15  sprint[-2]##the day finally arrived when i was sure i 'd leave sprint .
16  sprint plan[-2],sprint customer service[-3]##after years with that carrier 's expensive
plans and horrible customer service , portability seemed heaven-sent .
17  ##i 'd always eyed the nokia phones and had heard decent things about t-mobile , so i gave
it a whirl .
18  size[+1][u], sturdy[+2]##here 's the brief synopsis : the phone is tiny , cute , feels
kind of " plastic-like " ( as if it might break ) , but seems pretty sturdy .
19  game[+2],pim[+2],radio[+2]##it has lots of little cute features , my favorite being the
games and the pim ( personal information manager -- i.e.
20  organizer ) , and the radio !
```

Figure 6. Tokenizing method

D. *Part of Speech Tagging*

After tokenizing the tweets, POS tagging is used for parsing the tweets. Tagging is the process of assigning a part of speech marker to each word in an input text. Because tags are generally also applied to punctuation, tokenization's are usually performed before, or as part of, the tagging process are shown in Figure 7.

```
1   [('**********', 'JJ'),
('*', 'NNP'), ('Annotated', 'VBN'), ('by', 'IN'), (':', ':'), ('Minqing', 'NNP'), ('Hu',
'NNP'), ('and', 'CC'), ('Bing', 'NNP'), ('Liu', 'NNP'), (',', ','), ('2004', 'CD'), ('.', '.')]
2   [('*', 'JJ'), ('Department', 'NNP'), ('of', 'IN'), ('Computer', 'NNP'), ('Sicence', 'NNP'),
('*', 'NNP'), ('University', 'NNP'), ('of', 'IN'), ('Illinois', 'NNP'), ('at', 'IN'),
('Chicago', 'NNP'), ('*', 'NNP'), ('*', 'NNP'), ('Product', 'NNP'), ('name', 'NN'), (':', ':'),
('Nokia', 'JJ'), ('6610', 'CD'), ('*', 'NNP'), ('Review', 'NNP'), ('Source', 'NNP'), (':',
':'), ('amazon.com', 'NN'), ('*', 'VBZ'), ('*', 'NNP'), ('See', 'NNP'), ('Readme.txt', 'NNP'),
('to', 'TO'), ('find', 'VB'), ('the', 'DT'), ('meaning', 'NN'), ('of', 'IN'), ('each', 'DT'),
('symbol', 'NN'), ('.', '.')]
3   [('**********', 'JJ'),
('[', 'NNP'), ('t', 'NN'), (']', 'NNP'), ('excellent', 'NN'), ('phone', 'NN'), (',', ','),
('excellent', 'JJ'), ('service', 'NN'), ('.', '.')]
4   [('#', '#'), ('#', '#'), ('i', 'NN'), ('am', 'VBP'), ('a', 'DT'), ('business', 'NN'),
('user', 'NN'), ('who', 'WP'), ('heavily', 'RB'), ('depend', 'VBP'), ('on', 'IN'), ('mobile',
'JJ'), ('service', 'NN'), ('.', '.')]
5   [('phone', 'NN'), ('[', 'NN'), ('+3', 'NNP'), (']', 'NNP'), (',', ','), ('work', 'NN'),
('[', 'NNP'), ('+2', 'NNP'), (']', 'NNP'), ('#', '#'), ('#', '#'), ('there', 'EX'), ('is',
'VBZ'), ('much', 'JJ'), ('which', 'WDT'), ('has', 'VBZ'), ('been', 'VBN'), ('said', 'VBD'),
('in', 'IN'), ('other', 'JJ'), ('reviews', 'NNS'), ('about', 'IN'), ('the', 'DT'), ('features',
'NNS'), ('of', 'IN'), ('this', 'DT'), ('phone', 'NN'), (',', ','), ('it', 'PRP'), ('is',
'VBZ'), ('a', 'DT'), ('great', 'JJ'), ('phone', 'NN'), (',', ','), ('mine', 'NN'), ('worked',
'VBD'), ('without', 'IN'), ('any', 'DT'), ('problems', 'NNS'), ('right', 'VBD'), ('out', 'IN'),
('of', 'IN'), ('the', 'DT'), ('box', 'NN'), ('.', '.')]
```

Figure 7. POS tagging

E. *Aspect Extraction*

In this paper, aspect is extracted as important features and rated by the reviewers. An aspect may be a single word or a phrase for a particular domain is identified through the training process. In order to extract the aspect, the noun and noun phrases of the tweets are need to be searched as shown in Figure 8. Table 3 displays the extraction of the aspects from the tweets and the number of aspects found in the tweets.

```
[('PHONE', 192), ('I', 122), ('PHONE [ +2 ]', 36), (']', 29), ('SERVICE', 18), ('T ]', 18),
('SPEAKERPHONE', 16), ('RADIO', 16), ('FEATURE', 14), ('PHONE [ +3 ]', 14), ('T-MOBILE', 14),
('QUALITY', 12), ('SIZE', 12), ('BATTERY LIFE', 11), ('RADIO [ +2 ]', 11), ('VOICE', 11),
('NOKIA', 11), ('SCREEN', 11), ('RECEPTION', 10), ('[ +2 ]', 10), ('CAMERA', 9), ('[ -2 ]', 9),
('TIME', 9), ('LOT', 9), ('+2 ]', 8), ('SIZE [ +2 ]', 8), ('THING', 8), ('SPEAKERPHONE [ +2 ]',
8), ('T610', 7), ('POCKET', 7), ('MENU', 7), ('WAY', 7), ('+1 ]', 6), ('-2 ]', 6), ('BATTERY',
6), ('SCREEN [ +2 ]', 6), ('VOLUME', 6), ('ACCESS', 6), ('EMAIL', 6), ('SANSUNG', 6),
('HEADSET', 5), ('USE', 5), ('WORLD', 5), ('COLOR SCREEN', 5), ('PHONE I', 5), ('SOUND
QUALITY', 5), ('PLAN', 5), ('COUNTRY', 5), ('SIZE [ +1 ]', 5), ('SPRINT', 5), ('BATTERY LIFE
[ +3 ]', 4), ('ANYTHING', 4), ('AMAZON', 4), ('T', 4), ('ANYONE', 4), ('PERSON', 4), ('MUSIC',
4), ('SIZE [ +2 ]', 4), ('PICTURE', 4), ('RADIO [ +3 ]', 4), ('END', 4), ('CSR', 4), ('GSM',
4), ('INTERNET', 4), ('STUFF', 4), ('DESIGN', 4), ('AREA', 4), ('BATTERY LIFE [ +2 ]', 4),
('FEATURE [ +2 ]', 4), ('VOICE DIALING', 4), ('[ +1 ]', 4), ('WEIGHT', 4), ('NUMBER', 4),
('MINE', 4), ('QUALITY [ +2 ]', 4), ('DIALING', 4), ('NETWORK', 4), ('-1 ]', 4), ('BATTERY [ +2
]', 4), ('I.E', 3), ('SUPPORT', 3), ('COMBINATION', 3), ('[ -1 ]', 3), ('WARRANTY', 3),
('BUTTON [ -2 ]', 3), ('CELL', 3), ('LAPTOP', 3), ('PROBLEM', 3), ('QUALITY [ +3 ]', 3), ('SONY
ERICSSON', 3), ('FAMILY PLAN', 3), ('VIBRATION', 3), ('MODEL', 3), ('CARRIER', 3), ('PORT', 3),
('+3 ]', 3), ('RECEPTION [ +3 ]', 3), ('VOLUME [ -2 ]', 3), ('CELL PHONE', 3), ('DAY', 3),
('WEB', 3), ('LINE', 3), ('REASON', 3), ('LIFE', 3), ('RINGTONE [ +1 ]', 3), ('CUSTOMER
SERVICE', 3), ('SERIES', 3), ('RING', 3), ('DEPARTMENT', 3), ('COMPUTER', 3), ('BENEFIT', 3),
('WORK', 3), ('BIT', 3), ('LACK', 3), ('FACT', 3), ('EUROPE', 3), ('WEEK', 3), ('CINGULAR', 3),
('CELLPHONE', 3), ('MENUS', 3), ('STORE', 3), ('HARDWARE', 3), ('RECEPTION [ +2 ]', 3), ('FM
RADIO', 3), ('MONTH', 2), ('SPEAKERPHONE [ +3 ]', 2), ('APPLICATION [ +1 ]', 2), ('CABLE', 2),
('EARPIECE', 2), ('STYLE', 2), ('YEAR', 2), ('CALL', 2), ('LIGHT', 2), ('PIECE', 2), ('WEIGHT [
+2 ] [', 2), ('PRODUCT', 2), ('DEAL', 2), ('GAME', 2), ('WRITING', 2), ('KEY', 2), ('T-ZONE [
-2 ]', 2), ('PHONEBOOK', 2), ('SONY', 2), ('WEIGHT [ +2 ]', 2), ('DOWNSIDE', 2), ('MONEY', 2),
```

Figure 8. Aspect extraction

TABLE 3
LIST OF ASPECT EXTRACTION

Aspect Extraction	No of aspects
Service	18
Speaker phone	16
Radio	16
Quality	12
Size	12
Battery Life	11
Radio	11
Voice	11
Screen	11
Camera	9

F. *Identifying Opinion Words and Their Orientation*

In this method, the opinion words are the words, which express their opinions towards aspects. In this step, the aspect related opinion words should be identified. The opinion words are mostly verbs, adverb, adjectives, adjective and adverb verb arrangements. For each opinion word, we need to identify its semantic orientation, which will be used to predict the semantic orientation of each opinion sentence. To get the contextual information of a sentence, negations should be handled appropriately. If the opinion word is in negative relation, then its priority score is reversed for negation handling purpose. Figure 9 shows the identification of opinion words polarity and Table 4 displays the list of some opinion words and its sentimental analysis respectively. Figure 10 displays the graphical analysis of aspects and its polarity reviews

```
PHONE :          Positive =>  24.47      Negative =>  12.46
I :              Positive =>  24.41      Negative =>  13.07
PHONE [ +2 ] :           Positive =>  27.65      Negative =>  9.97
] :              Positive =>  24.46      Negative =>  12.95
SERVICE :                Positive =>  23.71      Negative =>  13.4
T ] :            Positive =>  24.34      Negative =>  12.98
SPEAKERPHONE :           Positive =>  28.57      Negative =>  3.9
RADIO :          Positive =>  28.0       Negative =>  12.0
FEATURE :                Positive =>  26.4       Negative =>  12.92
PHONE [ +3 ] :           Positive =>  30.09      Negative =>  9.73
T-MOBILE :               Positive =>  21.2       Negative =>  13.59
QUALITY :                Positive =>  19.66      Negative =>  11.11
SIZE :           Positive =>  22.61      Negative =>  12.17
BATTERY LIFE :           Positive =>  20.9       Negative =>  14.93
RADIO [ +2 ] :           Positive =>  27.27      Negative =>  9.09
VOICE :          Positive =>  17.57      Negative =>  10.81
NOKIA :          Positive =>  22.35      Negative =>  6.47
SCREEN :                 Positive =>  28.4       Negative =>  8.64
RECEPTION :              Positive =>  24.44      Negative =>  12.22
[ +2 ] :                 Positive =>  28.03      Negative =>  10.4
CAMERA :                 Positive =>  26.09      Negative =>  2.17
[ -2 ] :                 Positive =>  22.78      Negative =>  22.78
TIME :           Positive =>  18.03      Negative =>  21.31
LOT :            Positive =>  23.68      Negative =>  21.05
+2 ] :           Positive =>  28.03      Negative =>  10.4
SIZE [ +2 ] :            Positive =>  23.61      Negative =>  12.5
```

Figure 9. Identifying the polarity of the opinion words

TABLE 4
LIST OF SOME OPINION WORDS AND ITS SENTIMENTAL ANALYSIS

Opinion words	Positive	Negative
Service	23.71	13.4
Speaker phone	28.57	3.9
Radio	28.0	12.0
Quality	19.66	11.11
Size	22.61	12.17
Battery Life	20.9	14.93
Radio	27.27	9.09
Voice	17.57	10.81
Screen	28.4	8.64
Camera	26.09	2.17

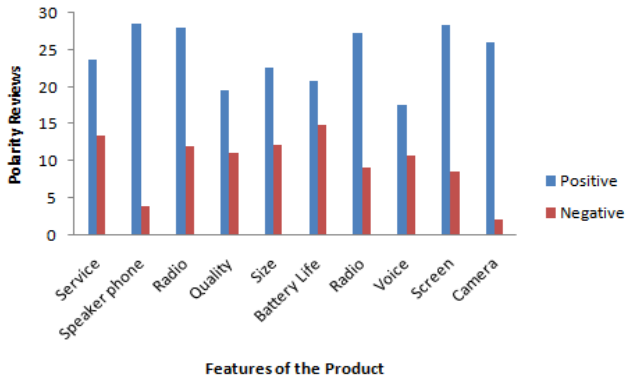


Figure 10. Graphical analysis of aspects and its polarity reviews

V. CONCLUSION

Opinion mining is a research domain dealing with automatic methods of detection and extraction of opinions and sentiments presented in a tweet. Opinion mining applications can result in creation of effective referral systems, financial analysis, market research and product development. In this paper, we implemented an idea to find sentimental analysis of the opinion words or phrases for each tweet from user reviews from Twitter in an efficient way. Proposed system successfully effective in defining the semantic polarity of input reviews. Aspect-based opinion mining aims to extract aspects and opinions from customer reviews in Twitter. This work, the aspects and their corresponding opinions not only help the consumers' decision-making process, but can also be used by manufacturers and new market researchers.

VI. REFERENCES

[1] Syed Akib Anwar Hridoy, M. Tahmid Ekram, Mohammad Samiul Islam, Faysal Ahmed and Rashedur M. Rahman." Localized twitter opinion mining using sentiment analysis", springer, 2015.

[2] Jantima Polpinij, Multilingual Sentiment Classification on Large Textual Data, IEEE Fourth International Conference on Big Data and Cloud Computing,2014,183-188.

[3] Samir, E.Abdelrahman,EbtsamAbdelhakamSayed,Reem Bahgat,The integration among disambiguation lexical resources for more effective phrase level contextual polarity recognition,Proceedings of 9th International Conference on Computer Engineering Systems(ICCES),2014,86-91

[4] Akshi Kumar, Teeja Mary Sebastian, " Sentiment Analysis on Twitter," IJCSI International Journal of Computer Science Issues, Vol 9, Issue 4, No 3, July 2012

[5] Su Su Htay and Khin Thidar Lynn," Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews", The Scientific World Journal Volume 2013 (2013), Article ID 394758, 5 pages

[6] L. Zhang and B. Liu, "Identifying noun product features that imply opinions," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '11), pp. 575–580, June

[7] Sonal Meenu Singh, Nidhi Mishra" Aspect based Opinion Mining for Mobile Phones", 2016 2nd International Conference on Next Generation Computing Technologies (NGCT-2016)

[8] M.Vadivukarassi, Dr.P.Aruna, N.Puviarasan,"Real Time Visualization Of tweets from Twitter Stream Analysis using Python", International Journal of Applied

Engineering Research(IJAER), ISSN 0973-4562,Vol. 11, No.4, pp: 490-500, 2016.(SCOPUS Indexed Journal)

[9] Jan Prechystal,"Mobile application's for customer review opinion", Procedia - Social and Behavioral Sciences 220 (2016) 373 – 381

[10] Ion Smeureanu, Cristian Bucur," Applying Supervised Opinion Mining Techniques on Online User Reviews", Informatica Economică vol. 16, no. 2/2012

[11] Sunidhi Dwivedi , Shama Parveen," Document Level Opinion Mining of Reviews of Mobile Phone Companies", International Journal of Innovative Research in Science, Engineering and Technology,Vol. 5, Issue 6, June 2016

[12] Ivan Cruz, Alexander Gelbukh, Grigori Sidorov," Implicit Aspect Indicator Extraction for Aspect-based Opinion Mining",, International Journal of Computational Linguistics and Applications, vol. 5, no. 2, 2014, pp. 135–152

[13] Shahab Saquib Sohaila, Jamshed Siddiquia, Rashid Ali," Feature extraction and analysis of onlinereviews for the recommendation of booksusing opinion mining technique", Perspectives in Science (2016) 8, 754—756