

languages[2]. The algorithm which performs opinion extraction, considers value of opinion holder whereas this paper takes the value of opinion holder as one. Generally sentimental word dictionaries will be used for labelling of small piece of data called “crunches”. These kinds of dictionaries contain certain threshold value for sentiment word and the defined value is used to decide sentiment of word is optimistic or pessimistic for subjective sentences. SentiWordNet V3.0 or WordNet are the online available sentiment word dictionaries [6]. For Example,

1.) Optimistic Sentiment in subjective sentence: “I like my new shoes” Defined sentence is expressed optimistic sentiment about the new shoes and we can decide that from the sentiment threshold value of word “like”. Threshold value of word “like” has optimistic numerical threshold value. Use this threshold value in the classification algorithm like naive-Bayes.

2.) Pessimistic sentiment in subjective sentences: “Two States is a flop movie” defined sentence is expressed pessimistic sentiment about the movie named “Two States” and we can decide that from the sentiment threshold value of word “flop”. Threshold value of word “flop” has pessimistic numerical threshold value. Use this threshold value in the classification algorithm like naive-Bayes.

3.) Neutral sentiment in subjective sentences: “I’m going for a party” defined sentence is expressed fact. It doesn’t carry any sentiment so we put this kind of statement in the neutral category. We can decide that the defined sentence is neutral because there is absence of words that express sentiment. Polarity, subjective detection and opinion identification all are very important things in this kind of sentiment analysis.

Another approach was Wilson’s [12] approach, which constituted the task of identification of context based polarity for a huge subset of sentiment expressions. This approach was called as phase-level sentiment analysis. Hu’s and Dave’s in their research primarily focused on extraction of opinion given the remarks. Hu’s [8] research featured a product aspect based research. It’s primary aim was to extract features of products and give product based summary. Kim and Hovy [3] in their first model analysed remarks for finding sentiments using word sentiment classifier with word net on a selected topic. Probability of sentiment words was used in the second model. The figure (Figure 2) gives a brief idea about the user interface.

Twitter Analytics - Sentiment Analysis and more

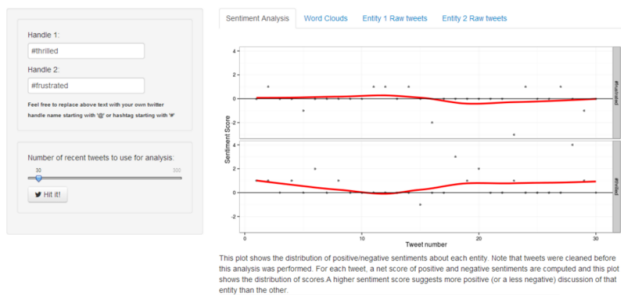


Figure 2: Sample User Interface

Sentiment Analysis for objective sentences is very trending research topic now-a-days because there are so many data origins which have objective sentences that carry sentiment but because of lack of proper algorithms and contexts we can't get the fruitful result from the objective sentences. According to recent article published by Ronen Feldman express that objective sentences that carry sentiment should be analysed for getting efficient sentiment analysis and this is one of the challenging task in sentiment analysis. [1], [10].

Origin of objective sentences consists of news, articles, blogs, social media etc. where we get good amount of objective sentences. [10] We consider following examples which are objective sentences but still carry sentiment. [1], [10], [11]

- “Windows keep crashing.” defines sentences carry pessimistic sentiment about Windows Operating System.
- “The headphones broke in two days.” defines sentence carry pessimistic sentiment about the earphones.
- “I feel relaxed after today’s session.” defines optimistic sentiment about person’s routine.

In this particular area only challenges have been recommended but still researchers are trying to find out efficient solutions to analyse these kinds of implicit opinions in the objective sentences. Available sentiment dictionaries don't have enough vocabulary to analyse objective sentences and categorise them efficiently into optimistic, pessimistic or neutral. Providing proper context or semantic orientation is also very important part of sentiment analysis of objective sentences.

III. RECOMMENDED CONCEPT

The paper recommends a sentiment analysis algorithm which uses collaborated opinion mining and has been described with the help of a case study. Algorithms and models have been recommended for analyzing sentiments and extracting opinions as shown in the survey. Identifying sentiment words first has been the primary concept among all models and algorithms. These words help find the place of the opinion present in the document. The opinions extracted, then are analyzed and polarity of opinion is discovered. This approach is the bottom-up approach, used in most of the algorithms [2]. A bottom-up approach is followed for the identification of opinions present. The recommended algorithm analyses the reviews/feedback given by customers or end-users word by word. Analysis of an opinion requires analysis of all words in a particular sentence. For this, the recommended algorithms make use of the thesaurus and word net [8, 10]. For the analysis of opinion, a database consisting of sentiment words has been used. In this database every word (i.e. sentiment word) is given a number. When the detection of this sentiment word happens, the number is saved in the database and is thereon used for evaluating the cumulative opinion number. Figure 2 shows the basic flow of the project.

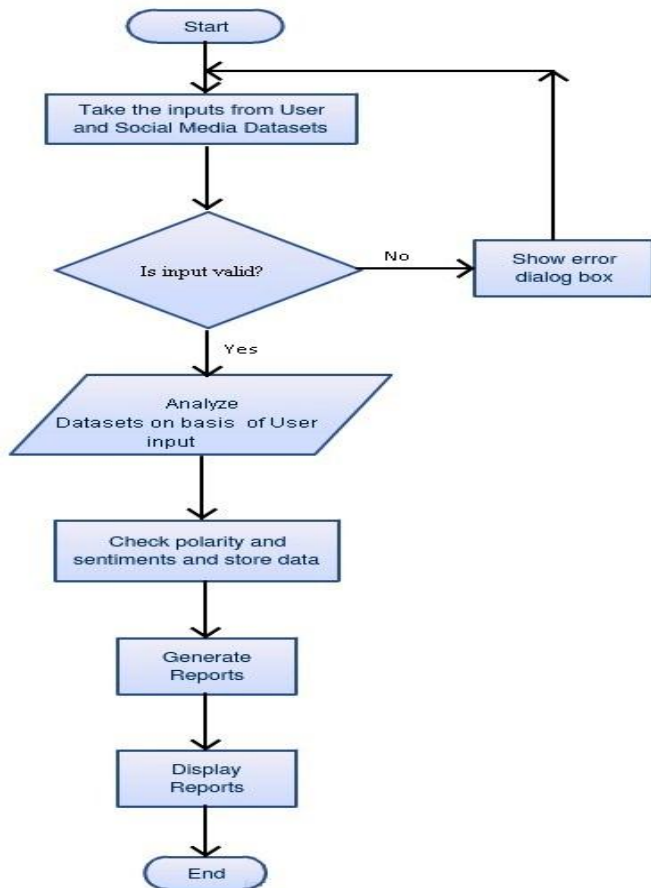


Figure 2: Flowchart

A. Recommended Algorithm

The algorithm we are to use is the baseline algorithm. Identification of the polarity of remarks is achieved by the algorithm discussed. In previous work there was no consideration of polarity of remarks in a sentence word by word. A case study has been used to explain the recommended work. Each and every remark or word is checked and compared by this algorithm to identify its polarity. The discussed algorithm will generate a numeric value (i.e. a number) for the opinion. If opinion number is high then the opinion is considered optimistic. A lower opinion number represents a pessimistic remark. This algorithm analyses the remarks word by word for each sentence [2]. All the sentiment words are identified and a combined number is given to each sentence. For identification of sentiment words, a database is maintained. An associated number for the opinion word is stored in the database. Each sentiment word is based on the strength or weakness of the sentiment used to assign a number to each sentiment word. The number ranges from zero to ten. Opinion has a higher number in the database if the sentiment word emotes strongly optimistic. Opinion has a lower number in the database if the sentiment word emotes strongly pessimistic. Opinion for each word is fetched from the database whenever there is match found for a sentiment word in a particular sentence. Estimation is made for the collaborated opinion number of that sentence. The number of the opinion score is decreased/ increased by a specific value if there is a negation in the sentence. The algorithm which is being used is as follows:

1. for every word.
2. Check whether it is a negation or sentiment word.
3. A number is assigned to every sentiment word (in the database). A number less than 5 represents pessimistic opinion (e.g. bad). A number greater than 5 represents optimistic opinion. How much low or high number can be decided using thesaurus.
4. If negation is present before sentiment is increased or decreased by 2 depending upon whether the sentiment number is high or low respectively.
5. For each comment or tweet provided by the user an average number is computed. The range of the number is from 0 to 10. The evaluation of collaborated opinion score is done as follows:
 If number < 2
 Very less
 If number > 2 but < 4.5
 Less
 If number > 4.5 but < 5.5
 Medium
 If number > 5.5 but < 8
 High
 If number > 8
 Very high

IV. IMPLEMENTATION

The implementation for the project includes four major steps: 1) Data Collection, 2) Data Pre-processing, 3) Polarity Analysis, and 4) Prediction. The data set to be used for this analysis is going to be the tweets concerning that to Air India. As specified before, the project is divided in two phases: that is, a polarity response analysis and sentiment analysis. Our area of concern or scope for this paper is limited to phase one only.

1. Data Collection: existing data set of twitter tweets and latest tweets via Twitter API. Data retrieval using keywords like '#AirIndia' and many more to be used.
2. Data Pre-processing: After collecting the "clean" data, it is transformed to the format we need. Lots of noisy, spam, irrelevant tweets would fill our dataset. Therefore we will convert the dataset to a format suitable as input for polarity analysis.
3. Polarity Analysis: A classifier is trained to classify the given tweets as: optimistic, pessimistic, neutral and irrelevant. Sentence (tweet) based analysis with a logistic regression classifier (Accuracy up to 80%).
4. Prediction: Statistics of the tweets labels is used to predict the Air India's reach and success.

The first step towards the implementation is the creation of word cloud which is a graphical representation of frequently used words in a collection of text files. The height of each word in this picture is an indication of frequency of occurrence of the word in the entire text. The word cloud is created by using a tool named R on the set of comments or tweets extracted by twitter or facebook for Air India. Word clouds are very useful for doing text analysis. The word cloud of our project is given in figure 1. The first step is the identification and creation of text files to turn into a cloud. These text files are stored in the location "/corpus/target". The second step is the creation of Corpus from the collection of text files i.e transforming text files into R- readable format. The third step is Data Processing on the text files in which symbols like "/" and "@" with a blank space and words like "a", "an", "the", "I", "He" are removed in order to remove skewness caused by the commonly occurring words. Also the whitespaces and punctuation are removed i.e cleaning of text files is done. The fourth step is creation of Structured data from the text file in which the entire corpus is converted into a structured dataset. The fifth step is the Creation of Word Cloud using the Structured form of the data.

V. CONCLUSION

In this paper we recommend an algorithm for calculation of value of collaborated opinion. On the basis of reviews given by people about Air India in numerous case studies, an average opinion value is calculated. The main idea here is to build a web application that would be used by the Air India employees/staff so as to get a response of all the Twitter tweets and comments of their Face book pages in a descriptive and graphical format. These results and observations would then be used to strengthen the promotional activities and campaigning of Air India for its worldwide growth. Also this analysis would provide them with sufficient insight on customer response so as to understand better how their customers feel about the services offered by Air India.

REFERENCES

- [1] Bing Liu, *Sentiment Analysis and Opinion Mining*, Morgan and Claypool Publishers, May 2012.p.1-19,27-28,44-45,47,90-101.
- [2] Lun-Wei Ku, Yu-Ting Liang and Hsin-Hsi Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora", 2006 American Association for Artificial Intelligence.
- [3] Soo-Min Kim and Eduard Hovy, "Determining the Sentiment of Opinions", conference of COLING, Geneva, 2004
- [4] Bo Pang and Lillian Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proceedings of ACL, 2004.
- [5] Online SentiWordNet dictionary origin
<http://sentiwordnet.isti.cnr.it/>.
- [6] Coursera Study origin on Sentiment Analysis
<https://class.coursera.org/nlp/lecture/145>.
- [7] M. Hu and B. Liu, "Mining and summarizing customer remarks," ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 168-177, 2004.
- [8] David Osimo and Francesco Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis", ICT Solutions for governance and policy modeling
- [9] Ronen Feldman, "Techniques and Application of Sentiment Analysis", Communication of ACM, April 2013, vol. 56.No.4.
- [10] Chihil Hung and Hao-kai Lin, "Using Objective Word in SentiWordNet to Improve Word-of-Mouth Sentiment Classification", IEEE Computer Society, P.47- 54, March-April 2013.
- [11] Theresa Wilson, Janyce Wiebe and Paul Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", Human Language Technology Conference and Empirical Methods in Natural Language Processing (HLT/EMNLP) Conference, pages 347-354, Vancouver, October 2005. c 2005 Association for Computational Linguistics.