International Journal For Technological Research In Engineering
Volume 5, Issue 7, March-2018 *ISSN (Online): 2347 - 4718*

STEP-BY-STEP TWITTER SENTIMENT ANALYSIS: VISUALISING TOP COLLEGES

Prof. B. Gopinathan[1], Akila .M[2], Gunjan Lall[3], Kaviyarasu .M[4]
[1]M.E.,(Ph.D)., Associate Professor, [2,3,4]Final Year Students,
Department Of CSE, Adhiyamaan College Of Engineering, Hosur

ABSTRACT: In today's era, time is exorbitant as estimated to money. Opinions and reviews are the most noteworthy hallmark in devising our views and persuading the success of a brand, product or service. Social Media has captured the attention of the entire world as it is thundering fast in sending thoughts across the globe, user friendly and free of cost requiring only a working internet connection. Amongst which twitter plays immense influence. Twitter is the most popular social media where people around the world voice out their opinions and reviews directly in the form of tweets. This paper performs a sentiment analysis of people's opinions(given in the form of tweets) regarding top colleges across India. In this busy world, people don't have enough time to look at each and every tweet and come to a conclusion about the colleges, which is a time consuming aspect. We are going to propose the system for detecting sentiment for dynamic tweets based on the threshold concept. The contingent essence of threshold value could be expected. Employing the threshold value, we are going to achieve accuracy in the project. By this people can come to a conclusion about the colleges briskly which inturn make the project time efficient.With the advent and growth of social media in the world, stakeholders often take to expressing their opinions on popular social media, namely twitter. While Twitter data is extremely informative, it presents a challenge for analysis because of its humongous and disorganized nature. This paper is a thorough effort to dive into the novel domain of performing sentiment analysis of people's opinions regarding top colleges in India. Besides taking additional preprocessing measures like the expansion of net lingo and removal of duplicate tweets.
Keywords: Twitter, sentiment analysis, net lingo

I. INTRODUCTION

Social Media has captured the attention of the entire world as it is thundering fast in sending thoughts across the globe, user friendly and free of cost requiring only a working internet connection. People are extensively using this platform to share their thoughts loud and clear. Twitter is one such well known micro-blogging site getting around 500 million tweets per day . Each user has a daily limit of 2,400 tweets and 140 characters per tweet. Twitter users post (or 'tweet') every day about various subjectslike products, services, day to day activities places, personalities etc. Hence, Twitter data is of Great germane as it can be used in various scenarios where companies or brands can utilize a direct connection to almost each of their client or user and thereby, improve upon their product. Consider a dis-satisfied costumer of a telecommunication company voicing out his/her grievances about a particular plan he/she is subscribed to. Twitter also serves as a huge platform for users to know more and get direct comments about a product or a service in which they are interested. Opinions and reviews in the form of tweets from customers, potential users and critics can easily influence the image and consequently, demand of a product/service being provide a company. Opinion is positive/negative about their offering becomes a crucial and pressing question for the organization to ask and monitor.

II. PROBLEM STATEMENT

The problem in sentiment analysis is classifying the polarity of a given text at the document sentence or feature/aspect level. Whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

ALGORITHM

An algorithm is an effective method that can be expressed within a finite amount of space and time and in a well-defined formal language for calculating a function. Starting from an initial state and initial input (perhaps empty), the instructions describe a computation that, when executed, proceeds through a finite number of well-defined successive states, eventually producing "output" and terminating at a final ending state.

One such used in this proposed system is Naïve Bayes. Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

Abstractly, naive Bayes is a conditional probability model: given a problem instance to be classified, represented by a vector {\displaystyle \mathbf {x} =(x_{1},\dots ,x_{n})} representing some *n* features (independent variables), it assigns to this instance probabilities:

{\displaystyle p(C_{k}\mid x_{1},\dots ,x_{n})\,}for each of *K* possible outcomes or *classes* {\displaystyle C_{k}}.

Rather than using Nielsen's word list as the prominent features for classification, chi-squared test was used to select the best k features for the polarity classification where k ranged from 10 to 15,000. The Naïve Bayes classifier utilizes all the features in these best k features and makes the 'naïve' assumption of independence of these features from each other.

III. EXISTING SYSTEM

In the existing system based on the initial expansion of the words theygoing to give sentiment process based on topics they adopted. Initially in the existing System iteration process is done (for example 1st tweets iteration they are going to take 100 tweets, within that 100 which words are coming with more positive or more negative count that words will be added as positive or negative before 2nd iteration). Here accuracy is less because after iteration immediately we considering positive or negative sentiment without considering left out words in tweets. The study of few existing system has been donewhich are listed below. Arora D., Li K.F. and Neville
S.W.,Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study, 29th IEEE International Conference on Advanced Information Networking and Applications, pp. 680-686, Gwangju, South Korea, March 2015
Sentiment analysis of the text data available on the web either in the form of blogs or at social media sites such as Twitter, Facebook, and LinkedIn, offers information through which to assess people's perspective of products and services that are of interest to them. Consumers routinely scour the Internet to assess other user's reviews for a product or a service before making their own decision. Likewise, these same data can potentially provide businesses a snapshot in time of the users' response to their products/services and even the trends over time. This information gained through sentiment analysis can then be used by businesses to make decisions about improving their products and services, and gain an increased edge over their competitors. In this paper, we illustrate the potential of sentiment analysis of Twitter data to gauge users' response to popular smart phone brands and their underlying operating systems. Specifically, our objective is to investigate whether the tweets available on the web are sufficient to gain useful insight about the performance of popular smart phone brands, their battery life, screen quality, and on the perceived performance of the
phones operating systems. Our results show that although the Twitter data does provide some information about users' sentiments to the popular smart phone brands and their underlying operating systems, the amount of data available for different brands varies significantly. This limitation makes the comprehensive analysis of users' response somewhat more challenging for some brands compared to others and consequently makes the comparison between brands almost impossible.
Kanakaraj M., Guddeti R M.R., Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques, 9th IEEE International Conference on Semantic Computing, pp.169-170, Anaheim, California, 2015
Mining opinions and analyzing sentiments from social network data help in various fields such as even prediction, analyzing overall mood of public on a particular social issue and so on. This paper involves analyzing the mood of including Natural Language Processing the society on a particular news from Twitter posts. The key idea of the paper is to increase the accuracy of classification by Techniques(NLP) especially semantics and Word Sense Disambiguation. The mined text information is subjected to Ensemble classification to analyze the sentiment. Ensemble classification involves combining the effect of various independent classifiers on a particular classification problem. Experiments conducted demonstrate that ensemble classifier traditional outperforms machine learning classifiers by 3-5%.
BahrainianS.-A,DengelA.,Sentiment Analysis and Summarization of Twitter Data",16th IEEE International Conference on Computational Science and Engineering, pp. 227-234, Sydney, Australia, December 2013
Sentiment Analysis (SA) and summarization has recently become the focus of many researchers, because analysis of online text is beneficial and demanded in many different applications. One such application is product-based sentiment summarization of multi-documents with the purpose of informing users about pros and cons of various products. This paper introduces a novel solution to target-oriented (i.e. aspect-based) sentiment summarization and SA of short informal texts with a main focus on Twitter posts known as "tweets". We compare different algorithms and methods for SA polarity detection and sentiment summarization. We show that our hybrid polarity detection
system not only outperforms the unigram state-of-the- art baseline, but also could be an advantage over other methods when used as a part of a sentiment summarization system.
Additionally, we illustrate that our SA and summarization system exhibits a high performance with various useful functionalities and features.
Neethu M.S. and Rajasree R., Sentiment Analysis in Twitter using Machine Learning Techniques, 4th IEEE International Conference on Computing, Communications and Networking Technologies, pp. 1-5, Tiruchengode, India, 2013
Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in source text. Social media is generating a vast amount of sentiment rich data in the form of tweets, status updates, blog posts etc. Sentiment analysis of this user generated data is very useful in knowing the opinion of the crowd. Twitter sentiment analysis is difficult
compared to general sentiment analysis due to the presence of slang words and misspellings. The maximum limit of characters that are allowed in Twitter is 140. Knowledge base approach and Machine learning approach are the two strategies used for analyzing sentiments from the text. In this paper, we try to analyze the twitter posts about electronic products like mobiles, laptops etc using Machine Learning approach. By doing sentiment analysis in a specific domain, it is possible to identify the effect of domain information in sentiment classification. We present a new feature vector for classifying the tweets as positive, negative and extract

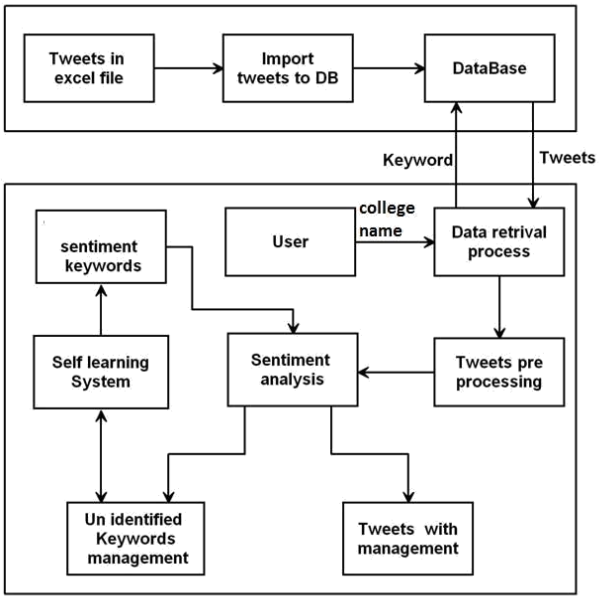
peoples' opinion about products.

IV. PROPOSED SYSTEM

We are going to propose the system for detecting sentiment for dynamic tweets based on the threshold concept. Based on the threshold value we are going to achieve accuracy in the project. In the proposed system initial expansion is done based on the topic selected. Based on the topic every word in the particular tweet, sentiment type of word is checked. Finally positive, negative or neutral count is incremented.

The left out word which is not in any sentiment type ,that word sentiment is decided based on the positive, negative and neutral count in that particular tweet. If positive count is more than negative and neutral them it will be considered positive sentiment only. Finally if left out word sentiment crosses threshold value, then that word sentiment is permanently considered.

SYSTEM ARCHITECTURE



Tweets Import Module

In this module, tweets are retrieved from the twitter API dynamically based on the college name input. To retrieve tweets from the twitter API account, first need to create twitter account in developer's console. After creating account we will get consumers token key and access token key, with the help of generated keys, we are going to communicate with twitter API to retrieve tweets.The retrieved tweets are imported into database.

Preprocessing Module

In this module, the tweets which are imported to database from the twitter API, these tweets consist of unnecessary words, whitespaces, hyperlinks and special characters.First we need to do filtering process by removing all unnecessary words, whitespaces, hyperlinks and special characters. Self Learning and word standardization System In this module, First we need to initialize the dictionary (first iteration dictionary).In the dictionary generally we need to initialize the positive, negative neutral and nouns. All big data and data mining projects based on the trained data, without trained data (initialization of words).So initialization of the trained data is very important. In the self learning system, we are doing word standardization ,here we are not considering past, present and future status of the words, only we are considering the word. Sentiment Analysis Module

In this module, preprocessed tweets are fetched from the database one by one. First we need check one by one keyword whether that keyword is noun are not, if noun we will remove it from the particular tweet. After that remaining keywords checked with sentiment type, whether that keywords are positive sentiment or negative sentiment or neutral sentiment. The remaining keywords in the tweet which does not belongs to any of the sentiment will be assigned temporary sentiment based on the more count of positive,negative and neutral.In the second iteration if the remaing:// support.word crosses the threshold of positive ,negative or neutral ,that keyword permanently added as expansion in the dictionary. Finally sentiment of the tweet is detected based on the positive, negative and neutral words in the particular tweet.

V. CONCLUSION

Sentiment analysis gives the optimal result depending on the threshold value generated.

In the project, opinions which are framed by the twitter users around the globe apropos whichever subject. Mechanization of formulating the opinions makes it easy to deal with colossal data lined up in the social websites such as Twitter on the basis of real events. Collating to the existing system, the proposed has a lot of innovations which includes lively fetching of the perspective from the social websites.

REFERENCE

[1] Arora D., Li K.F. and Neville S.W., Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study, 29th IEEE International Conference on Advanced Information Networking and Applications, pp. 680-686, Gwangju, South Korea, March 2015

[2] Choi C., Lee J., Park G., Na J. and Cho W., Voice of customer analysis for internet shopping malls, International Journal of Smart Home: IJSH,vol. 7, no. 5, pp. 291-304, September 2013.

[3] India'sBestColleges,2015, http://indiatoday. intoday.in/bestcolleges/2015/

[4] Kanakaraj M., Guddeti R M.R., Performance Analysisof Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques, 9th IEEE International Conference on Semantic Computing, pp. 169-170, Anaheim, California, 2015.

[5] King R. A., Racherla P. and Bush V.D., What We Know and Don't Know about Online Word-of-Mouth:A Review and Synthesis of the Literature, Journal of Interactive Marketing, vol. 28, issue 3, pp. 167-183, August 2014.

[6] Ministry of Human Resource Development, http:// mhrd.gov.in/ statist
[7.] Posting a tweet,https://support.twitter. com/ articles/15367-posting-a-tweet.
[8] Twitter usage /Company Facts, https:// about. twitter.com/ company

