ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Sentimental Analysis of User Attributes using Social Network Graph

Monica Shatani, Abhishek Kunden, Sabah Memon, Gershwin Fernandes

B.E. Student, Dept. of Computer, Trinity College of Engineering & Research, Pune, India

ABSTRACT: Positivity and negativity of a person are two important attributes of emotions and spirits. Emotions although being an important factor in individual's everyday life are many times forgotten in the development of systems to be used by individuals. Sentimental analysis focuses on understanding the subjective information such as good emotions, good moods, and good spirits. Sentimental analysis is useful in social media to monitor or to characterize the overall mood or opinion towards the specific subject or object and get the viewpoint through the positive or negative statements on web. Our proposed methodology is implemented with the input sets obtained from the user's interaction on social media (Twitter) and feeding them to a natural language model for observation and processing. Also the system implemented checks the validity of the result which is obtained by comparing the results obtained from the natural language processing model. In order to go from data to information, to knowledge and to wisdom we need to reduce the complexity of data. Complexity can be reduced using cluster analysis. Audio to text conversion which is the highlight of our implementation is enforced using a set of attributes viz acoustic model, dictionary model and language model. The main aim is to distinguish the users with positive emotions and the users with negative emotions which will provide the best recommendation to overcome negativity.

KEYWORDS: emotions; spirits; moods; sentimental analysis; clustering; natural language model

I. INTRODUCTION

Consumers are day by day looking more on feedbacks posted on the social network channels to make a variety of decisions feeding from what movies to watch to what business to invest in.Various researches in the past has suggested that consumers follow these reviews and consider them more equitable and transparent than the traditional sources. Since social network websites have become popular media for people to interact, share their opinions, enterprises have sought the opportunities to rank this data for business intelligence applications like online analytical processing and enterprise marketing services.

Sentimental analysis is the study of using a machine to determine the polarity of an idea-whether it is positive, negative or neutral. It is the form of Natural language processing for tracking the mood of the public about a particular brand or topic.

There are several challenges in sentimental analysis:

i. The first is an opinion word that is considered to be positive in one situation may be negative in other situation.
ii. A second challenge is that people don't always express their views in the same way.

In Sentimental analysis, however, "the picture was great" is distinct from the "the picture was not great". Every Individual have their own opinions and hence these sentences are inconsistent . Most reviews will have both the positive and negative comments, which is somewhat manageable by evaluating the sentences one at a time.However,in the more informal medium like twitter or blogs, People are more likely to have different point of view expressed in the same sentence which is easy for human to understand butmore difficult for computer to parse. Sometimes it becomes difficult in understanding what someone anticipated based on a short piece of text because of the lack of context.For example, "That movie was good as its latest movie" this dependents on an individual what he has got to express based on the previous model.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

The main aim is to distinguish the users with positive emotions and the users with negative emotions which will provide the best recommendation to overcome negativity. Audio to text conversion which is the highlight of our implementation is enforced using a set of attributes viz acoustic model, dictionary model, language model.

i. **Acoustic model**: Models the relationship between audio signal and the vocal units in the language.
ii. **Language model**: Responsible for modeling the word sequences in the language.
iii. **Dictionary model**: For grammatical relations between words in a sentence.

Complexity of data can be reduced using Clustering analysis. For analyzing the text and sentiments we have used Stanford CoreNLP the next section we discuss why we chose Stanford CoreNLP.

II. WHY DID WE CHOOSE STANFORD CORENLP

Stanford core NLP is a natural language analysis tool. It includes many NLP tools some of which are the part-of-speech (POS) tagger, the parser, the coreference intent system, opinion mining, reboot pattern learning, and the open information extraction tools.

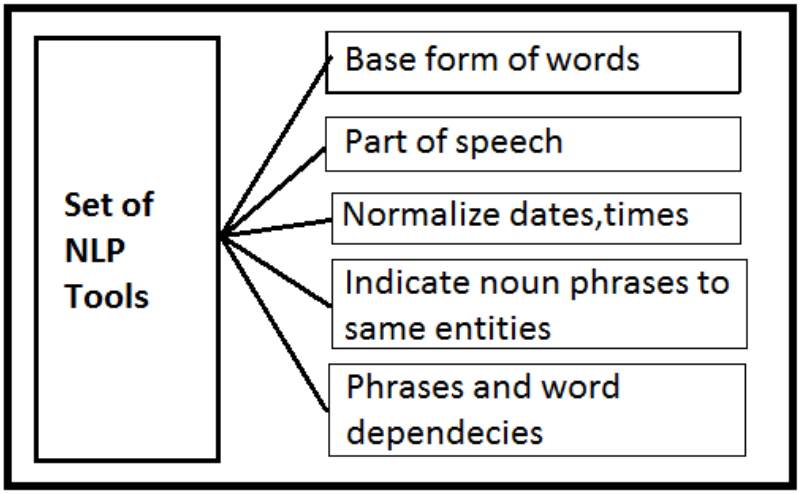

Fig 1. NLP Tools

There are some reasons why we chose to use Stanford NLP:

i. A toolkit of grammatical analysis tools.
ii. Fast, steady analysis of arbitrary texts.
iii. Good quality text analytics.
iv. Support for a number of major languages.
v. Interfaces available for different major modern programming languages.
vi. Capacity to run as a simple web service.

III. RELATED WORK

In [5] authors describe the design and use of the Stanford CoreNLP toolkit. The usage of the Stanford CoreNLP system (an annotation-based NLP processing pipeline) has been clearly illustrated here. One of the main aims of writing the paper as it can be clearly seen was to make the usage of the framework easier for the masses rather than making it complex by adding functionality. While [5] focused on speech to text conversion using an NLP framework from Stanford, [4] gives one a clearer idea of the further processing of textual data for sentiment analysis. Now-a-days most of the data flowing in and out of the Internet comes from social networking sites, moreover from the posts and statuses and comments that are updated by the second. This paper demonstrates the use and function of a classifier for

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

rule-based classification, supervised learning and machine learning. A complementary approach has been incorporated where each classifier could contribute to achieve a potential result as compared to other classifiers. In[10] sentiment extraction and analysis leads to drawing conclusions regarding the exact sentiments of the person who posted what he did which enables various consumer service executives to provide the appropriate solution to an issue faced by the customer. The same logic could be used to analyse sentiments to increase the level of positivity of the content posted online. Traditional text mining focused on analysis of facts; this drawback led to the adoption of sentimental analysis which concentrates on attitudes. Opinion mining and sentiment analysis go hand in hand differentiating the positive, negative and neutral posts which in turn could be used for reducing negativity over the Internet and also for promoting positive behaviour clubbed with sarcasm detection.

IV. SYSTEM DESIGN

Our Proposed system is implemented with the input sets obtained from the user's interaction on social media (Twitter) and feeding them to a natural language model for observation and processing.

There are three approaches:

i. Private Login
ii. Public Login
iii. Audio upload

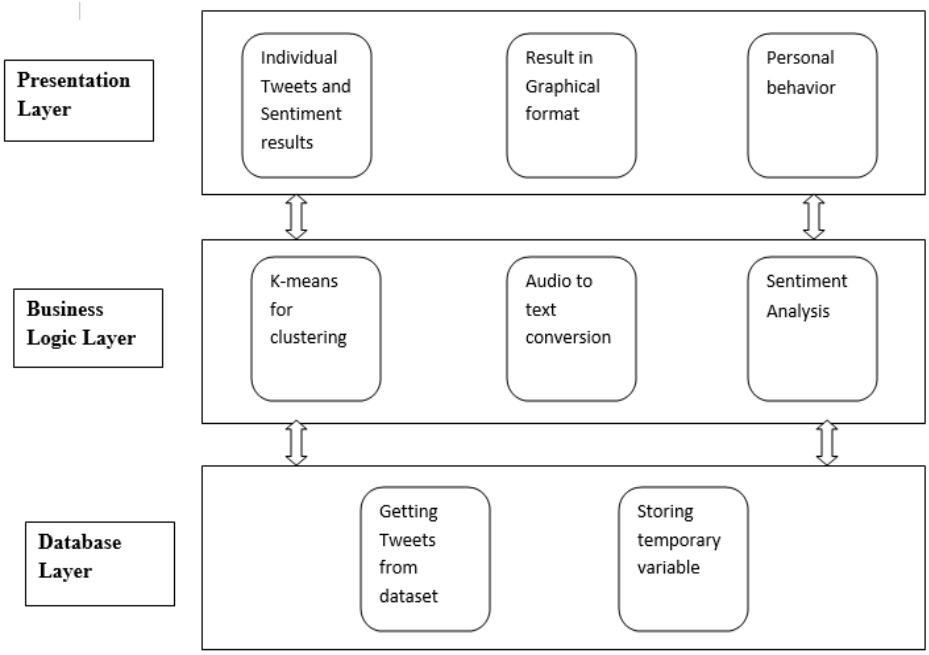

Fig 2. System Design

Firstly, we have private login for individual tweets which will be processed to filter the useful information and then it gives a score to the tweets. The positive and negative scores reflect the opinion of user's interaction. Secondly we have public tweets i.e. twitter dataset which have multiple tweets of multiple end user which are analysed and then gives a score for those tweets. Lastly, we have an audio upload. To analyse the data, Audio is converted to text using sphinx library and natural language processing is done.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Complexity of data to be handled using cluster analysis. Classifiers are good, very good, neutral, bad, and very bad. With the help of clustering we can then generate a graph and with this graph we can now find the number of people who are in the particular range on twitter with the help of the tweets on a particular topic or generally too. We can use this further for the learning systems.

V. PROPOSED METHODOLOGY

Categorization of the system

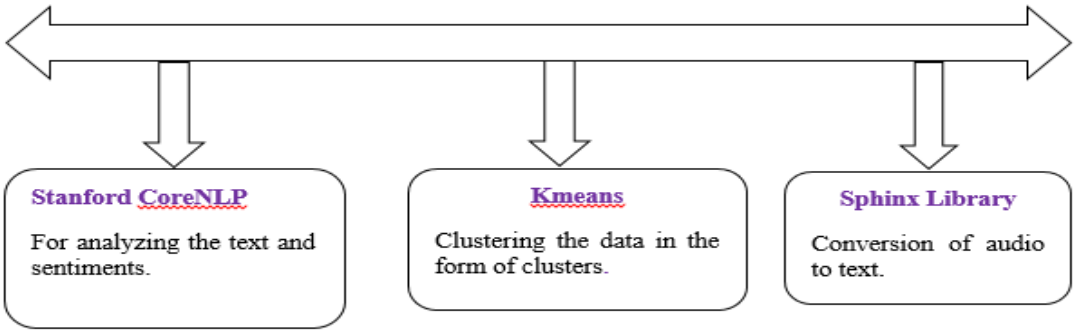

Fig 3. Categorization of System

A) Stanford CoreNLP

Following is the execution flow:

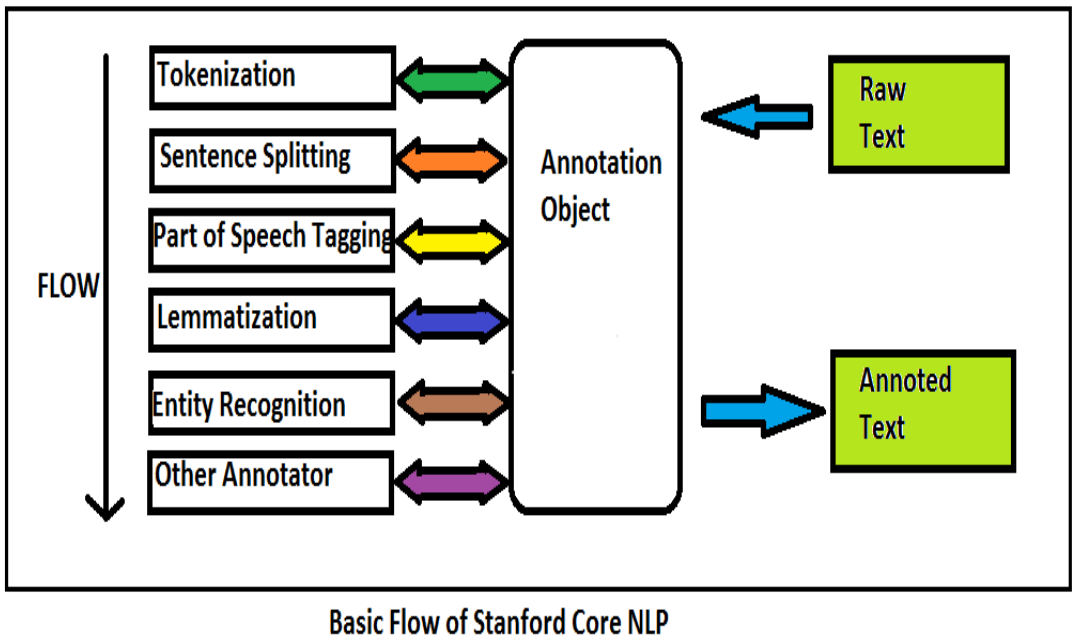

Fig 4. Execution Flow of Stanford CoreNLP

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Step 1: **Tokenizer:**Divides text into sequence of token which corresponds to words.
Step 2: **SentenceSplitting:** Split a sequence of token into sentences. Uses a decision tree to decide its end of sentence or not when facing a "."
Step 3: **Lemmatization: Finds** the correct dictionary word form.
Step 4: **Named Entity Recognition:** Recognizes named (person, location), numerical (number, percent) and temporal (date, time) entities.
Step 5: **DependencyParsing:** Provides a Representation of grammatical relations between words in a sentence.
Step 6: **Coreference: This** occurs when two or more Expressions in text refers to same person or thing.

B) K-Means

Let X = x1,x2,x3,. . . . . ..,xn be the set of data points
V = v1,v2,. . . . . .,vc be the set of centres.
Step 1**:** Randomly select `c' cluster centres.
Step 2: Calculate the distance between each data point and cluster centres.
Step 3: Assign the data point to the cluster centre whose distance from the cluster centre is least possible of all the cluster centres...
Step 4: Recalculate the new cluster centre, using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

Where, `c_i' represents the number of data points in the cluster.
Step 5:Recalculate the distance between each data point and new obtained cluster centres.
Step 6: If no data point was reassigned then stop, otherwise repeat from step3).

C) Sphinx Library

Sphinx is a pure Java speech recognition library. It converts the speech recordings into text with the help of CMUSphinx Phonic models. It helps to identify speakers, adapt models, and align existing transcription to audio for timestamping.
Step 1:Add the jar files into the dependencies of your project.
Step 2:There are several high level recognition interfaces.

i. Live Speech Recognizer
ii. Stream Speech Recognizer
iii. Speech Aligner

Here, we are using Stream Speech Recognizer for the interfacing.
Step 3: Set the attributes which are:

i. Acoustic model
ii. Dictionary model
iii. Language model

Step 4: These attributes are then passed to recognizer using configuration object.
Configuration is used to supply appropriate and alternative attributes to recognizer.
Step 5:Configuration object sets the path to
The required attributes.

```
configuration.setAcousticModelPath ("file:"+pathString+"\\WebContent\\lib\\en-us");
configuration.setDictionaryPath ("resource:/edu/cmu/sphinx/models/en-us/cmudict-en-us.dict");
configuration.setLanguageModelPath ("file:"+pathString+"\\WebContent\\lib\\cmusphinx-5.0-en-us.lm.dmp");
```

Step 6: StreamSpeechRecognizer uses Input stream as the speech source.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

VI. SIMULATION RESULTS

Categorization done into five groups:

i. Positive >=3

ii. Very Positive >3

iii. Neutral = 2

iv. Negative <=1

v. Very Negative<1

The result and analysis of data processed using the methodology explained above gives us the following observations:

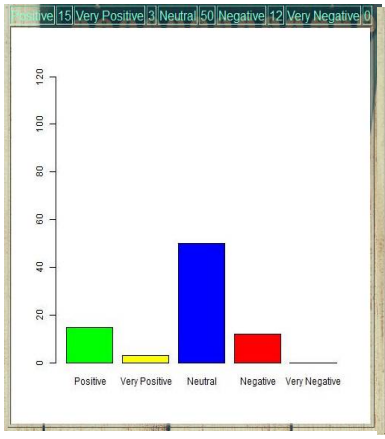

Fig 5. Private Tweets

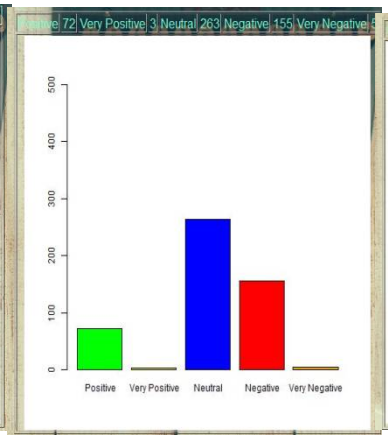

Fig 6. Public Tweets

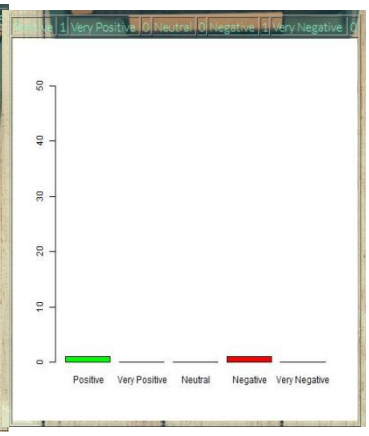

Fig 7. Audio Dataset

This analysis has categorized people in five groups from very negative to very positive. Twitter's tweets are thus used as a major tool for creating a social awareness and emotional support. Thus, when the followers or followees tweet a positive or negative text or audio dataset, we aim at analysing the emotions behind it. This will allow Twitter users to understand the level of positivity or negativity attracted by their page.

REFERENCES

1. Nitish Bhardwaj, Anupam Shukla, Pradip Swarnakar "Users' Sentiment Analysis in Social Media Context using Natural Language Processing" (Department of Information Technology, ABV-Indian Institute of Information Technology and Management, Gwalior,MP, India)ISBN:978-0-9891305-6-1 ©2014 SDIWC
2. Jianbo Yuan, Quanzeng You and Jiebo Luo"Sentiment Analysis Using Social Multimedia"© Springer International Publishing Switzerland 2015 A.K. Baughman et al. (eds.), Multimedia Data Mining and Analytics, DOI 10.1007/978-3-319-14998-1_2
3. Sudipta Roy, Sourish Dhar, Arnab Paul, Saprativa Bhattacharjee, Anirban Das, Deepjyoti Choudhury"CURRENT TRENDS OF OPINION MINING AND SENTIMENT ANALYSIS IN SOCIAL NETWORKS" IJRET: International Journal of Research in Engineering and Technology,Volume: 02 Special Issue: 02 | Dec-2013
4. Rudy Prabowo, Mike Thelwall "Sentiment Analysis: A Combined Approach"School of Computing and Information Technology University of Wolverhampton Wulfruna Street WV1 1SB Wolverhampton,UK
5. ChristopherD.Manning,MihaiSurdeanu,JohnBauer,JennyFinkel,StevenJ.Bethard,DavidMcClosky"TheStanfordCoreNLPNaturalLanguage ProcessingToolkit"
6. http://nlp.stanford.edu/software/patternslearning.html
7. http://nlp.stanford.edu/software/tokenizer.html
8. http://nlp.stanford.edu/software/classifier.html
9. http://nlp.stanford.edu/software/relationExtractor.html
10. G.Vinodhini*,RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012