Sentiment Analysis Based on Comments from Online Social Network

Vedant Patil, Jayesh Thakur, Kapildev Yadav and Prof. Deepti Lawand

Department of Information Technology, PCE, Navi Mumbai, India - 410206

Abstract— Internet is the platform where most of us share our happiness or other feelings. Recent years are devoted in studying and mining the data which is on social platform. This task includes understanding explicit and implicit information conveyed by sentiments. It can be extracted from the comments on social media using dictionary-based sentiment analysis or Review-Seer. Comments of the person are important to analyze the sentiments of the person at the time of writing the comment. The task is to classify the comments into positive, negative and neutral sentiments further into different emotions, for which it uses the concept of Plutchik's wheel of emotions and further makes a dictionary. The system will take input from user to classify and predict the emotions and strength of that emotion (Negative Emotions). There are basic eight emotions and system will primarily focus on negative emotions. Plutchik's wheel of emotion gives joy and sadness, anger and fear, trust and disgust, surprise and anticipation. The use of Plutchik's wheel of emotions will provide the real emotional view of comments. The confidence of the will be given which will indicate the strength of feeling. It uses fuzzy logic approach using Naïve Bayes or decision tree algorithm for prediction and generates output.

Keywords—Sentiment Analysis, Plutchik's Wheel, Machine Learning, Data Mining.

1. Introduction

Recently there has been a growing interest in social media and using it to update lifestyle. These comments entered by user contains pure emotions that needs to be extracted using different data mining algorithms. The task of mining sentiments and opinions from natural language is difficult one. It involves an intense understanding of most of the implicit and explicit information which is conveyed by structure of language. The availability of a dynamic corpus contains the user generated data, such as reviews for products or polling data. Big data is the large amount of easily available data on web, Social media, remote sensing data, etc. in form of structured data, semi-structured or unstructured data. We can use this large data for sentiment analysis. Sentiment analysis is the opinion mining used on the web for identifying the text. It is nothing but to get the real voice of people for specific product, services, movies, news, issues from online social networking site like Twitter. This data contains many important aspects which will be helpful in judging the turn of tide in market trend.

2. Literature Survey

A. Fine-grained Sentiment Analysis with 32 Dimensions[1].: This system does deal with range of total 32 emotions. It uses concept of Plutchik's wheel of emotion to classify comments into different 32 sentiments. The mathematical model of Naïve Bayes is used for classification and prediction uses intensity-based technique. Using Naïve Bayes classifier to classify into 32 emotions also makes increase in accuracy. The eight basic emotions given is used as classes to classify emotions into.

B. Rule-based Emotion Detection on Social Media: Putting Tweets on Plutchik's Wheel[2]: This paper studies and analyze sentiments beyond polarity and uses Plutchik's wheel of emotions. It uses extension of Rule based emission model. This model thinks beyond the normal metrics of sentiment analysis using polarity and uses Rule-Based Emission Model (RBEM) algorithm (Tromp and Pechenizkiy, 2013) that can be used for polarity detection assigning new messages a label that is one of positive, neutral, negative. Important in algorithm is that positivity and negativity are opposite and allows negation to simply invert the emission. RBEM uses pattern matching and uses wildcards for it. The model used is compact as well as complete which works well with RBEM-Emo which is stated as extension of Rule Based detection algorithm.

C. Combining Strengths, Emotions and Polarities for Boosting Twitter Sentiment Analysis[3]: This paper proposes an approach for boosting twitter sentiment classification using different sentiment dimensions as meta-level features. This research shows the combination of sentiments improves the twitter sentiment classification tasks. The scopes of tweets are categorized upon some categories as polarity, emotion, strength. It does different testing with different types of algorithm. It uses classification approach like OpinionFinder Lexicon, AFFIN Lexicon, SentiWordNet Lexicon, SentiStrenght Lexicon, Senti140 method, NRC Lexicon. So, when it

classifies tweets into polarity classes, we are essentially projecting these multiple dimensions to one single categorical dimension. But also, sentiment classification of tweets can lead to loss of valuable sentiment information.

D. Sentiment Analysis on Product Review Using Plutchik Wheel of Emotion with fuzzy logic[4]: The model consists of stemming and stopword techniques. This filtering removes almost all unwanted noise from comment. The filtered comment is then split to get the separate words for comparing. Then each single word is compared with the sentiment words dictionary. If the word is matched with the positive or negative dictionary then it is placed in the corresponding box, that is positive word in positive words text and in the same way negative words are placed. The comparison is done between number of positive word and number of negative words in a given comment. The condition is checked whether the positive words are more or negative and accordingly the comment is decided to be positive or negative. If both the positive and negative words are same or if there are no positive or negative, the comment is treated as neutral comment.

2.1 Summary of Related Work

Sentiment Analysis has been largely developed in the recent years due to the requirement of sentiments of people. Due to this, a large number of importance has been given to neural networks, machine learning, etc. For ex, deep neural network classifiers has been proposed in [1] where in the model proposed is including all the 32 emotions and has used the naïve bayes technique which we implementing in our model. The emotions are further put on the Plutchik's and the emotion is derived. Existing.

System Has Following Steps:
Data collection using twitter API: Publicly large sets of Twitter data are not available. Hence, they first extracted twitter data through twitter API.
Data Preprocessing[4]: It involved cleaning of data by spell correction punctuation etc. Reducing noise from the data.
Applying Classification Algorithm: The Classification Algorithm is applied on tweets to categories them with highest accuracy.
Classified tweets and result: The tweets are further classified three defined categories. Result of which is displayed in form of pie chart.

One of the techniques used in sentiment analysis is Rule Based Emotion Detection (RBEM) [2]. The RBEM thus efficiently increases the exactness of the developed model where the accuracy is guaranteed and is tightly coupled with the plutchik's wheel of emotion. The data collection, preprocessing, clustering, sentiment classification, prediction[4] are various processes that are been applied to get accurate output with the help of plutchik's wheel of emotion.

3. Proposed Work

The proposed system will primarily classify emotion into positive, negative, neutral and further into 8 basic emotions. Firstly, the data required for analysis will be divided into testing and training datasets, this dataset are downloaded from official twitter APIs. The Naive Bayes classifier will be trained according to this dataset.

3.1 System Architecture

The system architecture is given in Figure 1.
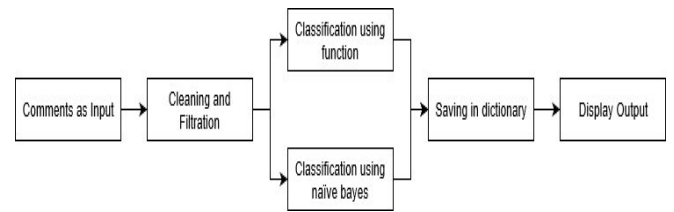

Fig. 1 Proposed system architecture

A. Comments as Input: The system will accept texual inputs which are in comments format. These comments will be entered by user after signing in on our website and typing comments on their page. These comments will be further cleaned to analyze them.

B. Cleaning And Filtration: The second part receives the input comments entered by user. This comment cannot be directly used to analyze the emotions as it contains some amount of excessive information which is not use full for sentiment extraction, such as Nouns, Names of places, etc. So we will be using extraction using predefined tokenizer known as TextBlob. It will extract the noun and sentiment defining word. Textblob can tag the word with part of speech which can be further useful while training naïve bayes classifier. Other processes are included like lemmatization.

C. Classification using Predefined Function: After cleaning input comment it is further used for extracting basic emotion. We will be using predefined function for extracting basic classification into positive negative and neutral. The classifier will be trained on training dataset.

D. Classification using naïve bayes: Naive Bayes classifier will train using previous training and test

datasets. This naive Bayes classifies the comments into 8 classes named as anticipation, joy, anger, sadness, surprise, disgust, fear and trust. This comment will be saved along with its tag of class into sentiment dictionary for further learning of system. We can train classifier with:

cl = NaiveBayesClassifier(train).

We can also find accuracy of test set using:

cl.accuracy(test)

After training output which has to be given using naive Bayes classification technique. We know the formula which can be used for classification as,
$P(A|B) = P(B|A) * P(A) / P(B)$
Where, $P(A|B)$ is probability the A belongs to class B, $P(B|A)$ is evidence, $P(A)$ is probability of class A is seen and similar with B.
The $P(A|B)$ is calculated for each word and then the class tag is selected with maximum probability. Maximum probability is selected and saved into dictionary for further increasing accuracy of classifier.

E. Saving to dictionary: After classification of comment into three basic emotion and then prediction into one of the eight emotion, comment is further saved into database for future predictions. The comment is saved along with tag of the emotion and further prediction will be done.

E. Output Prediction: After saving into dictionary, output will be displayed to user. Output will give the basic classified emotion of that comment. It will give positive, negative or neutral along with polarity of comment. Also it will give the prediction of emotion of comments from eight basic emotion given in the plutchik's wheel of emotion.

3 Requirement Analysis
The implementation detail is given in this section.

3.1 Software
Software requirements are Html and Bootstrap supporting browser, as the systems user interface is based on that versions. Also Flask framework is used for classifier.

3.2 Hardware

Strong internet connection is recommended for website access. Otherwise no hardware requirements are there.

3.3 Dataset and Parameters
[1]Dataset used contains comment along with tags of positive, negative and neutral emotion. Used dataset mostly have no neutral comment as it doesn't have any effect on classifier performance.

REFERENCES
1. *Xianchao Wu, Hang Tong, Momo Klyen,* "Fine grained Sentiment Analysis with 32 Dimensions", A.I. & Research Microsoft Development Co. Ltd., Department of Mechano-Informatics, The University of Tokyo.
2. *Erik Tromp, Mykola Pechenizkiy*, "Rule-based Emotion Detection on Social Media: Putting Tweets on Plutchik's Wheel", The Netherlands, 15 Dec 2015.
3. *Felipe Bravo-Marquez, Marcelo Mendoza, Barbara Poblete,* "Combining strength, emotions and polarities for boosting Twitter sentiment Analysis", Chile, 11 Aug 2013.
4. *Dhanshri Chafale, Amit Pimpalkar,* "Sentiment Analysis on Product reviews using Plutchik's Wheel of Emotion with Fuzzy Logic", Nagpur University, Dec 2015
5. UCI Machine Learning Repository: Sentiment Labelled Sentences Data Set. Available: *https://archive.ics.uci.edu/ml/datasets/Sentiment -Labelled-Sentences* [Accessed:21Aug-2018].