International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 11
November 2017

Emotion Detection of Tweets using Naïve Bayes Classifier

Hema Krishnan
Research scholar, School of Engineering, CUSAT
M. Sudheep Elayidom
Associate Professor, School of Engineering, CUSAT
T. Santhanakrishnan
Scientist E, NPOL

Abstract

Emotion plays an important role in human life. Emotion is a summary of the number of tweets distributed across different emotions such as fear, joy, sadness etc. There is increasing body of research on understanding the human emotions over the last quarter-century,. Twitter is a large and rapidly growing micro blogging social networking website where people express their opinions in a short and simple manner of expressions. It is a common practice that merchants selling products on the Web ask their customers to review the products. In twitter number of customer reviews on different products is appearing. Electronic products are a common domain in which number of customer reviews appears. Among the electronic products, mobile phones are a major domain in which customers have a keen interest. Customers express their emotions on tweets. Naïve Bayes classifier is a simple probabilistic model that works well on text classification. By using Twitter content, it classifies the different emotions like joy, sadness, anger, love, fear, thankfulness, and surprise for the twitter content. This paper focuses on emotion recognition from tweets using Naïve Bayes classifier. A NoSQL database MongoDB is also used for efficient storage of tweets.

Keywords: Emotion, Naïve Bayes, MongoDB, Tweets

Introduction

Emotions play an important role in social media and they are used to stimulate cognitive processes for strategies making. Emotions represent 1000 another form of language universally spoken and understood. Identification and classification of emotions has been a research area since Charles Darwin's age. Textual information can be collected from many sources, such as books, newspapers, web pages, e-mail messages, etc.

Nowadays Internet is the most popular communication medium also rich in emotion. Emotions can be extracted from textual input by analyzing punctuation, emotional keywords, syntactic structure and semantic information with the help of natural language processing techniques. There is a large amount of textual information available in the web. It is interesting to extract emotions for different goals like those of business. For example, in luxury goods, the emotional aspects as brand, uniqueness and prestige for purchasing decisions, are more important than rational aspects such as technical, functional or price. In this case customer is happy to buy a product even with high prices. Emotional Marketing aims to stimulate emotions in customer for tying him to brand and so increase the sell of product/service. Nowadays it isn't the product to be sold, since for each category there is a wide choice, but the focus is the relationship that the consumer establishes with the brand and with the emotions which the product communicates.

Twitter is a large and rapidly growing micro blogging social networking website where people express their opinions in a short and simple manner of expressions. It is a common practice that merchants selling products on the Web ask their customers to review the products. In twitter number of customer reviews on different products is appearing. Electronic products are a common domain in which number of customer reviews appears. Among the electronic products, mobile phones are a major domain in which customers have a keen interest. This makes it difficult for a potential customer to read them in order to make a decision on whether to buy the product.

An emotion is a particular feeling that characterizes a state of mind, such as joy, anger, love, fear and so on. Automatic emotion detection from text has attracted growing attention due to its potentially useful applications. For examples, psychologists can better assist their patients by analyzing their session transcripts for any subtle emotions; reliable emotion detection can help develop powerful human-computer interaction devices; and deep emotional analysis of public data such as tweets and blogs could reveal interesting insights into human nature and behavior. Emotion is expressed as joy, sadness, anger, surprise, hate, fear and so on. According to W. Gerrod Parrot [8], the human emotions are classified through an emotion hierarchy into six classes at primary level which are Love, Joy, Anger, Sadness, Fear and Surprise. Naïve Bayes classifier is an efficient classifier in terms of accuracy. This paper aims to automatically detect the emotions from tweets using Naïve Bayes classifier. Tweets are extracted and stored in a NoSQL database MongoDB [11] which stores the data in json format. Section 2 details the related works, section 3 and section 4 explains the modular design and Naïve Bayes classifier respectively. Section 5 gives the experimental result and final section 6 gives the conclusion.

Related works

S. Radha Krishna et.al [1] studied different spectral features such as MFCC, pitch chroma, skewness and centroid for emotion recognition. The emotions considered in this study are Fear, Anger, Neutral, and Happy. The system is evaluated for various combinations of spectral features. It is established that the combination of MFCC and skewness gave better recognition performance when compared with other combinations. These experiments are conducted and evaluated using Gaussian Mixture models (GMMs).The data base used in this study is Telugu emotion speech corpus (IIIT-KGP).

Taner Danisman et.al [2] studied, automatic classification of anger, disgust, fear, joy and sad emotions in text. The study was conducted on ISEAR (International Survey on Emotion Antecedents and Reactions) dataset. For the classification they have used Vector Space Model with a total of 801 news headlines provided by "Affective Task" in SemEval 2007 workshop which focuses on classification of emotions and valences in text. They have compared their results with ConceptNet and powerful text based classifiers including Naive Bayes and Support Vector Machines. Their experiments showed that VSM classification gives better performance than ConceptNet, Naive Bayes and SVM based classifiers for emotion detection in sentences. An overall F-measure value of 32.22% and kappa value of 0.18 for five class emotional text classification on SemEval dataset which is better than Navie Bayes (28.52%), SVM (28.6%) was obtained.

Swati D. Bhutekar et.al [3] presents a methodology to extract emotion from the text at real time and add the expression to the textual contents during speech synthesis. This paper also focuses on implementation of creation of Corpus, emotion recognition module etc. In text analysis, all emotional keywords and emotion modification words are manually defined. The test was carried out on set of textual sentences and preliminary rules written for 34 different emotions. These rules are used in an automated procedure that assigns emotional state values to words. These values are then used by speech synthesizer to add emotions to speech & input sentence. Pitch detection algorithm has been implemented for pitch recognition. The system is language dependent.

Changqin Quan et.al[4] , make an analysis on sentence emotion based on emotion words using Ren-CECps (a Chinese emotion corpus). Some classification methods (including C4.5 decision tree, SVM, NaiveBayes, ZEROR, and DecisionTable) have been compared. Then a supervised machine learning method (Polynomial kernel method) is proposed to recognize the eight basic emotions (Expect, Joy, Love, Surprise, Anxiety, Sorrow, Angry and Hate). Using Ren-CECps, we get the emotion lexicons for the eight basic emotions. Polynomial kernel (PK) method is used to compute the similarities between sentences and the eight emotion lexicons. Then the experiential knowledge derived from Ren-CECps is used to recognize whether the eight emotion categories are present in a sentence. The experiments showed promising results.

Ali Houjeij, Layla Hamieh et.al [5] designed a system that adopts a novel approach for emotional classification from human dialogue based on text and speech context. Their main objective was to boost the accuracy of speech emotional classification by accounting for the features extracted from the spoken text. The

proposed system concatenates text and speech features and feeds them as one input to the classifier. The work builds on past research on music mood classification based on the combination of lyrics and audio features. The innovation in our approach is in the specific application of text and speech fusion for emotion classification and in the choice of features, Furthermore, in the absence of benchmark data, a dataset of movie quotes was developed for testing of emotional classification and future benchmarking. The comparison of the results obtained in each case shows that the hybrid text-speech approach achieves better accuracy than speech or text mining alone.

Amira F. El Gohary et.a l[6] are concerned with the automatic detection of emotions in Arabic text. This construction is based on a moderate sized Arabic emotion lexicon used to annotate Arabic children stories for the six basic emotions: Joy, Fear, Sadness, Anger, Disgust, and Surprise. Their approach achieves 65% accuracy for emotion detection in Arabic text.

Shadi Shaheen, Wassim El-Hajj et.al k [7], proposed a framework for emotion classification in English sentences where emotions are treated as generalized concepts extracted from the sentences. They generated an intermediate emotional data representation of a given input sentence based on its syntactic and semantic structure. They then generalized this representation using various ontologies such as WordNet and ConceptNet, which resulted in an emotion seed called an emotion recognition rule (ERR). Finally, a suite of classifiers are used to compare the generated ERR with a set of reference ERRs extracted from a training set in a similar fashion. The used classifiers are k-nearest neighbors (KNN) with handcrafted similarity measure, Point Mutual Information (PMI), and PMI with Information Retrieval (PMI-IR). When applied on different datasets, the proposed approach significantly outperformed the existing state-of-the art machine learning and rule-based classifiers with an average F-Score of 84%.

Naïve Bayes Classifier

Naïve Bayes classifier [9] is based on the so - called Bayesian theorem with the naïve assumption of independence between every pair of features. This classifier in spite of the apparently over-simplified assumptions has worked quite well in many real- world situations. It is very fast and has a good performance, better in some cases than more sophisticated methods. The Naïve Bayes model with Gaussian is equivalent to a mixture of Gaussians (GMM) with diagonal covariance matrices. The main advantages of this classifier are the conditional independence assumption, which helps to obtain a quick classification, and the probabilistic hypotheses (results obtained as probabilities of belonging of each class).

Bayesian classification is based on Bayes' theorem. Bayes' theorem is named after Thomas Bayes, a nonconformist English clergyman who did early work in probability and decision theory during the 18th century. Let **X** be a data tuple. P(H/**X**) is the posterior probability, or a posteriori probability, of H conditioned on **X**. In contrast, P(H) is the prior probability, or a priori probability, of H. Similarly, P(**X**/H) is the posterior probability of **X** conditioned on H. $P(\boldsymbol{X})$ is the prior probability of ***X*.** Bayes' theorem is useful in that it provides a way of calculating the posterior probability, $P(H/\boldsymbol{X})$, from $P(H)$, $P(\boldsymbol{X}/H)$, and $P(\boldsymbol{X})$.

Bayes' theorem is

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}.$$

System design

The proposed methodology can be divided into different modules. The first module fetches the tweets. Second module extracts and stores the tweets in MongoDB [11] Nosql database and third module extracts emotions from tweets using Naïve Bayes Classifier.

A. Fetching tweets

In this module, tweets are fetched from the twitter website. The package twitteR is imported for the same. There are two sub modules within this module.

1) Twitter OAuth

Here the application should authenticate with twitter so as to access the twitter data. For this, an account has to be created in twitter. Four parameters [10] will be provided by twitter for the developers using which the developers and their applications can be authenticated. The four parameters are: consumer key, consumer secret, access token and access secret. After authentication, data can be fetched from twitter.

2) Extraction of tweets

Tweets are extracted from the site using searchTwitter() function. The function takes the following inputs:

- Search query to be issued to twitter.
- Maximum number of tweets to be returned.
- Language.
- Date since which the tweets have to be obtained.

The search for tweets is done using the hashtag of respective phone name. The search can also be done using other twitter handles as well.

B. Processing and storing tweets

In this module, tweets after processing are stored into the database. Mongodb [11] is used as the database. This module can be viewed as two submodules [12].

1) Processing of tweets

Here the tweets are subjected to a set of procedures which makes them suitable for sentiment analysis in the next module. The following steps are involved in the processing of tweets:

• Removal of re-tweets.

• Removal of handles like # etc.

• Removal of punctuations.

• Removal of digits.

• Removal of web addresses like https:// etc.

• Removal of blank spaces.

• Convert the entire text to lower case.

2) Storing tweets

The processed tweets are then converted into a dataframe and inserted into a mongodb [11] collection. MongoDB database is a NoSQL document oriented database which stores the data in JSON format.

C. Extraction of emotions from MongoDB

Emotions are extracted from tweets stored in MongoDB [11] using Naive Bayes Classifier. In this paper we are considering emotion categories like fear, joy, sadness etc. Tweets from the database are fed to the classifier and it classifies the emotions into different emotion classes.

Experiments and Results

The methodology is applied to latest 500 tweets extracted from Twitter. Five mobile phone brands are considered for study- Samsung, nexus, iphone, lenovo, motorola. Interface is created using shiny. Emotion categories for Samsung brand is shown in figure 1.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 11
November 2017

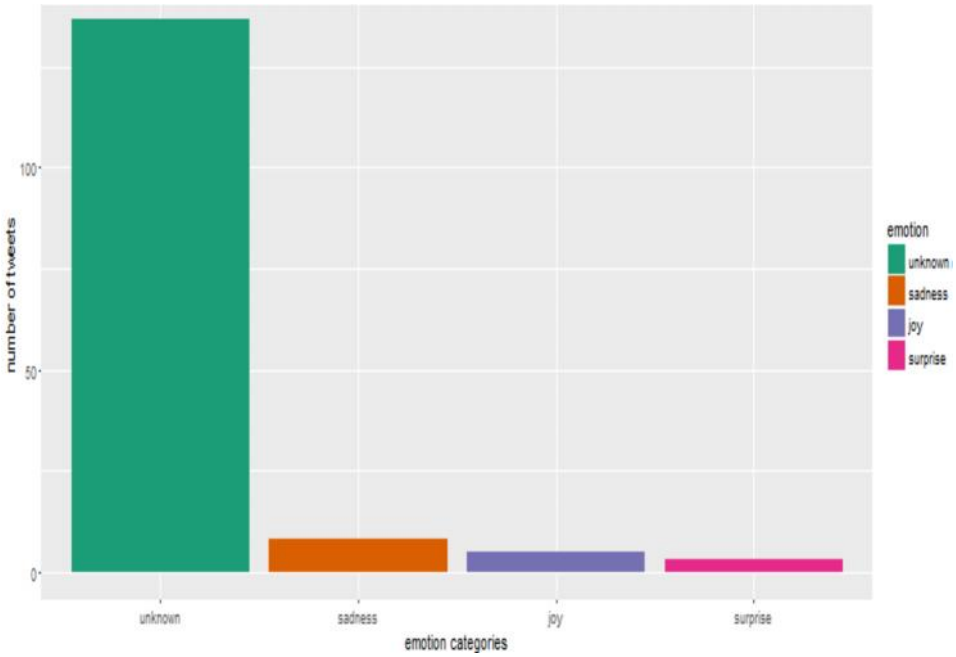

Figure 1: Emotion categories

Evaluation metrics used are precision, recall and accuracy.

Precision is the measurement of correctness.

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive}+\text{False Positive})}$$

Recall is the measurement of completeness.

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive}+\text{False Negative})}$$

Accuracy is overall performance of individual classifier is measured by:

$$\text{Accuracy} = \frac{\text{No of correctly labeled tweets}}{\text{Total no of Tweets in the dataset}}$$

Table 1 depicts the precision, recall and accuracy corresponding to the emotions sad, joy and surprise extracted from latest 500 tweets.

TABLE 1
EVALUATION METRICS

Emotion	Precision	Recall	Accuracy
Sad	1	0.982	99.31
Joy	1	0.98	99.31
Surprise	0.8	1	99.31

Conclusion

In this paper, we have introduced a methodology for automatically extracting emotion from tweets related to mobile phones. We have considered five mobile phone brands and latest 500 tweets are extracted. For storing tweets we are using a NoSQL database MongoDB and the tweets are stored in JSON format. Since Naive Bayes is a supervised learning algorithm, it leads better results when compared to unsupervised methods. The use of MongoDB allows efficient storage and retrieval of tweets in json format.

References

1. **S. Radha Krishna, R. Rajeswara Rao, "**Automatic Text-Independent Emotion Recognition Using Spectral Features", in Journal of Innovation in Computer Science and Engineering, Vol. 5(1), July – Dec 2015@ ISSN 2278-0947.
2. **Taner Danisman, Adil Alpkocak, "**Feeler: Emotion Classification of Text Using Vector Space Model",
3. Swati D. Bhutekar, Prof. M. B. Chandak, "Corpus Based Emotion Extraction to implement prosody feature in Speech Synthesis Systems", in International Journal of Computer & Electronics Research, Volume 1, Issue 2, August 2012, ISSN : 2778- 5795.
4. Changqin Quan, Fuji Ren, "**Sentence Emotion Analysis and Recognition Based on Emotion Words Using Ren-CECps", in** International Journal of Advanced Intelligence Volume 2, Number 1, pp.105-117, July, 2010.
5. Ali Houjeij, Layla Hamieh, Nader Mehdi, Hazem Hajj, " **A Novel Approach for Emotion Classification based on Fusion of Text and Speech", in 19th International Conference on Telecommunications.**
6. Amira F. El Gohary, Torky I. Sultan, "A Computational Approach for Analyzing and Detecting Emotions in Arabic Text", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 3, Issue 3, May-Jun 2013, pp.100-107.
7. Shadi Shaheen, Wassim El-Hajj et.al, "Emotion Recognition from Text Based on Automatically Generated Rules", in 2014 IEEE International Conference on Data Mining Workshop.
8. Parrott, W.G, "Emotions in Social Psychology," in Psychology Press, Philadelphia 2001.
9. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Second Edition.
10. https://sites.google.com/site/miningtwitter/questions/sentiment/.
11. http://www.tutorialspoint.com/mongodb/.
12. Hema Krishnan, M. Sudheep Elayidom, T.Santhanakrishnan, " Sentiment Analysis of tweets for inferring popularity of mobile phones", in International Journal of Computer Applications, Jan 2017 edition.