Recognizing Emotion in Customer's Feedback Mail Using Naive Bayes Approach

[1]Namitha A R, [2]Shwetha S N, [3]Sunil Kumar B R, [4]Shashikala S V

[1,2]Assistant professor, [3]M.tech, [4]Prof. & HOD

Department of computer science

BGSIT Mandya, India

Abstract-*this paper present a text processing technique for identifying the emotion present in the email. Users share their opinions, sentiment and also complaints about the product in the form of review through mails to the concerned person in the company. The process of emotion recognition in these customer reviews on product is one of the difficult task. It is also difficult for the mail reviewer to classify the mails based on emotion and prioritize the response accordingly. In this paper, we proposed efficient technique for detection of emotion in customer reviews on products.*

I. INTRODUCTION

Emotion is a particular type of feeling that represents various states of mind such as joy, fear, anger, love and so on. Emotion Detection from text is an important component of Artificial Intelligence; it plays a key role in human-computer interaction. A person can express emotions by speech, facial expressions, hand gestures and written text. Sufficient amount of work has been done in speech and facial emotion recognition but recognition of emotions through text still needs attraction of researchers. From applicative point of view, detecting the human emotions in text is becoming increasingly important in computational linguistics.

On the web there is huge amount of textual information which is quite interesting to extract emotions from these textual information for special purpose such as business. For example, in luxurious products available online, the emotional aspects like brand, uniqueness and prestige for purchasing decisions are weighted more by customer as compared to rational aspects such as technical, functional and cost. If a customer is emotionally satisfied he can purchase a product at high price. Emotional Marketing targets the customer's emotions to encourage him to opt for a particular brand and so results in increase of product/service sales.

Nowadays there is a wide range of products available, but the main target is to create confidence in customer about a product/service he communicates. There are various emotional models used to build emotion recognition system. One of the latest model suggested is the hourglass model which has been inspired biologically and psychologically-motivated, and is based on the idea that the emotional states results from the selective activation or deactivation of different resources in brain. Another very common model used in Ekman's model which divides emotions into six universal categories.

Several current approaches associated with emotion detection are based on methods of supervised learning, where a huge set of annotated data (the emotions are labeled as text) is needed to train the model.

II. LITERATURE SURVEY

The literature survey discusses topics relating to emotion classification and description about emotions.

A. Emotion

Emotion is a strong feeling which derives from one's method of forming judgment about circumstances, event or relationships with others .Emotions are complex. The study of emotion in psychology started in the 1970s and till date many different theories have being proposed, studied and scrutinized. Many dimensional models of emotion have been studied, researched and developed, but only a few amongst them remain as the dominating models. Some of these dimensional models can be used for emotion classification in text, which can be document, sentence, short message or tweets. The models are used for data collection in emotion classification.

For example the Circumflex model is a 2D model which was developed based on 8 emotions categories and 28 emotion words were placed in these 8 categories and the Positive Activation and Negative Activation model (PANA) is yet another model which classifies emotion words on the scale from high positive activation to low positive activation, these two models can be used for data collection for emotion analysis task. But models like Lovheims cube cannot be used for data collection or analysis because emotion in that is defined on levels of monoamine transmitters from the brain and we cannot detect levels of these transmitters from text. Most of the research work carried out in the field of emotion classification mainly uses Ekman's model for classification of emotions. Paul Ekman carried out a study to identify emotion based on different facial expressions between different cultural people, this study was done to find out whether people from different cultures have same facial expressions to represent certain set of basic emotions. Ekman's model provides six discrete emotion categories namely happiness, sadness, anger, disgust fear and surprise.

B. Emotion Classification

Emotion classification is a task wherein the aim is to detect and recognize types of feelings through the expression of texts, such as anger, disgust, fear, happiness, sadness, and surprise. Emotion detection may have different applications such as finding out how happy the citizens are, providing better services to an individual and suggestions in helping an individual who is in anxiety which can be identified through the outgoing texts and emails. Many authors have done noteworthy work in this field. Liza Wikarsa and Sherly Novianti Thahir developed a text mining application to detect emotions of Twitter users that are classified into six emotions, namely happiness, sadness, anger, disgust, fear and surprise. Preprocessing, processing and validation are the three main phases of text mining that were used in this application.

Tasks such as case folding, cleansing, stop-word removal, emoticons conversion, negation conversion, and tokenization of the training data as well as the test data were conducted in the preprocessing phase. Weighting and classification using Naive Bayes algorithm was done in the processing phase and a tenfold cross validation was carried out in the validation phase for measuring the accuracy generated by the application. This model obtained 83% accuracy for 105 tweets. The authors focused on social media data rather than text documents and also emphasized on the need of proper preprocessing steps to obtain useful results. Li Yu, Zhifan Yang, Peng Nie, Xue Zhao and Ying Zhang tackles the task of multi-source emotion tagging for online news. A new classification model is proposed with two layer logistic regression, this approach takes output from a basic classifiers and combines them in a new classifier, hence providing a more accurate prediction. This research was considered as an initial step towards multi-source tagging, the results can further be improved by using emotion dictionary and feature selection. Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, Amit P. Sheth proposed the idea of overcoming a bottleneck of lack of coverage of emotional situations in datasets used for emotion identification tasks. They experimented with various feature combinations and also with the effect of the size of the training data.

III. PROPOSED SYSTEM

Emotion detection application is used to discover customer’s feelings on product through mails. Emotion detection software will predict the customer emotions from the mail that may involves seven types of feelings. Emotions of Twitter's users which can be classified into six emotions, particularly anger, disgust, fear, happy, sad, and surprise. Three important phases of the email emotion detection utilized on this application had been text collection, preprocessing, and processing. Activities carried out within the preprocessing section had been case folding, cleaning, stopword removal, emoticons conversion, negation conversion, and tokenization to the learning information and the test data established on the sentiment analysis that carried out morphological evaluation to construct a number of models. Within the processing section, it performed weighting and classification utilizing the Naive Bayes algorithm.

Emotion detection application makes use of Naive Bayesian methods which is used to foretell the customer feelings. It could actually extract the data from live mail which is unstructured, tremendous and dynamic. To organize the accrued knowledge into pre-outlined categories that can be used for performing text analysis by way of preprocessing. Assemble the compatible units centered on the information set through processing then validate the emotions of mails within the information set.

A. EMOTIONS

The categorization of emotions has often been studied from two principal techniques: basic emotions and core influence.

- *Basic Emotions*: Basic emotion theorists think that people have a small set of normal feelings that are discrete. More than a few researchers have attempted to establish a number of general emotions which are universal amongst all people and vary one from an additional in important ways. A trendy example is a go-cultural study of 1972 by means of Paul Ekman and his colleagues, where they concluded that the six common emotions are anger, disgust, fear, happy, sad, and surprise.
- *Core Affect Model*: Core influence model of emotion characterizes human feelings by defining their positions along two or three dimensions. That's, most dimensional units incorporate valence and arousal dimensions.
- *Emotion Analysis in Text*: Effort for emotion evaluation on Twitter knowledge entire by Bollen and his colleagues. They tried to find a relationship between overall public mood and social, fiscal and other principal pursuits. They extracted six dimensions of mood (anxiety, depression, anger, vigor, fa- tigue, confusion) utilizing a multiplied variant of POMS (pro- file of temper States), a psychometric instrument. They located that social, political, cultural and fiscal pursuits have a enormous, and immediate outcomes on the various dimensions of public mood.

B. STOPWORDS

The Discontinue phrases are on the whole probably the most ordinary phrases together with articles (a, an, the), auxiliary verbs (be, am, is, are), prepositions (in, on, of, at), conjunctions (and, or, nor, when, even as), and it record together with bad verbs (now not, is just not, does now not, don't, must now not, and many others.), auxiliary verbs (be, am, is, are), prepositions (in, on, of, at).In addition, we changed the phrase —very! with blank and the word —clean no longer clean! is replaced via —clean not!. That do not provide additional growth for engines like google however broaden the computational complexity by means of growing the size of the dictionary.

Example:

For instance, —i'm happy.ll => —i'm happy.ll => —i'm happy.ll

—i'm not very happy.ll => —i'm not happy.ll => —i'mNOThappy.l

In this example the phrase —happyl and —not happyl is used to create new words —happyl and —NOThappyl. This means, we are able to discriminate the phrase —happyl having positive meaning and it is classified as the word belongs to happy class. In the identical means, the new phrase —NOThappyl has a negative meaning and it is classified as the word belongs to sad class.

IV. SYSTEM ARCHITECTURE

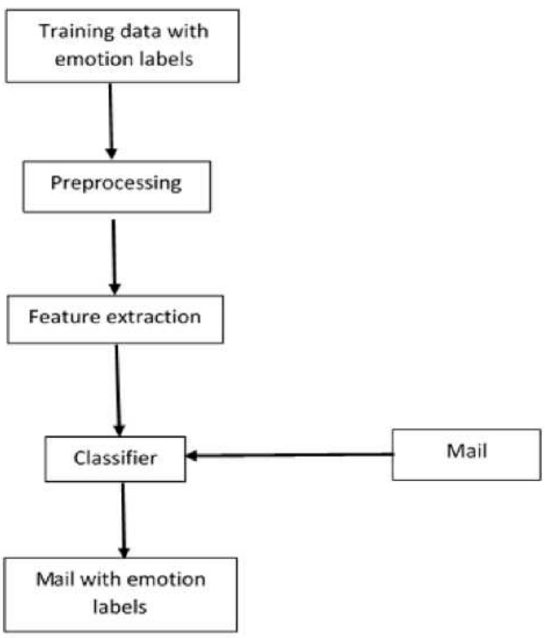

Figure:1 Architecture of emotion classification process

A. Training data with emotion tables

Dataset Collection The dataset was obtained from. For creating the dataset, first six different emotion words for six emotion categories were collected from existing psychology theory. The collected dataset was then divided into testing and training sets. The training sets were used to train the classifiers so that the test data is correctly labeled.

B. Preprocessing

Preprocessing of the collected data is of utmost importance because the mails will have several words and hence there is a presence of slang, URLs, user-mentions which do not contribute in any manner to the classification process, in fact the presence of such elements can mislead the classifier. The preprocessing steps include the following:

- Lower casing all the words.
- Remove all URLs and other non-useful words.
- Replacing letters and punctuations that are repeated more than twice with two same letters (Eg.happpy->happy).

C. Feature Extraction

Feature Extraction After data preprocessing, the stopwords are removed from the mail by using a stopword file, this is done because they usually constitute large components of sentences and they do not provide useful information. After stopword removal the feature extraction process is done. In this paper bag of words is considered for training both the classifiers that are used.

D. Classifier

Naive Bayes is for classification purpose because Naive Bayes works effectively for text classification.

V. CONCLUSION

The objective of this work is to develop a methodology for automatic detection of emotions from text. To train the model customer review Data has been collected from internet for different emotion categories. In this work various data preprocessing step has been applied to remove noise in the data. Classifications of different emotion words present in the sentences are used to find the category of emotion classes.

REFERENCES

[1] D.Farber—Twitter Hits 400 Million Tweets Per day,Mostly Mobile,l06thjune2012.[online].Available:www.c net.com/news/twitter-hits-400-million-tweets- per-day-mostly mobile/.[Accessed 10th September 2015].

[2] RizwanaIrfan, Christine K. King, Daniel Grages, Sam Ewen, Samee U. Khan,Sajjada. Madani, Joanna Kolodziej, Lizhe Wang, Dan Chen, Amma R Rayes,NikolaosTziritas, Cheng - ZhongXu, Albert Y. Zomaya, Ahmed SaeedAlzahrani,AndHongxiang Li —A Survey on Text Mining in Social Networks, ll The Knowledge Engineering Review, United Kingdom, (2004) pp.1-24.

[3] Susan Dumais John Platt David Heckerman —Inductive Learning Algorithms and Representations for Text

Categorization!, Published by ACM, 1998.

[4] Ananthi.J —A Survey Web Content Mining Methods and Applications for Information Extraction from Online

Shopping Sitesll, International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, pp. 4091-4094.

[5] R. Xin Chen, MihaelaVorvoreanu, and Krishna Madhavan —Mining Social Media Data for Understanding Students' Learning Experiences IEEE transactions on learning technologies, manuscript id 1, (2013), pp. 1-14.