

# Sentiment Analysis to Improve Emotional Health of User

Pranav Rane

Dept. of Computer Engineering  
D. J. Sanghvi College of  
Engineering  
Mumbai, India

Kashyap Bhansali

Dept. of Computer Engineering  
D. J. Sanghvi College of  
Engineering  
Mumbai, India

Prof. Sindhu Nair

Dept. of Computer Engineering  
D. J. Sanghvi College of  
Engineering  
Mumbai, India

## ABSTRACT

The primary idea of the project is to change the mind-set of the user by providing various kinds of media or activities with the help of Sentiment Analysis. The user's current mind-set is detected using sentiment analysis and the negative emotion is countered by showing appropriate media to the user. The input will be the emotional state of user in words and will be used to determine the exact emotion. This will allow us to show tailor-made results that are capable of overpowering the currently negative mind-set. The media response will change from user to user and will vary at different times of the day. A combination of fine sentiment analysis and intelligent profile management will be used to accomplish the results. A model is designed to learn from the user's behavioral traits. Thus the goal of transforming the user's mental state to achieve longer positive mentality shall be achieved.

## General Terms

Machine Learning, Emotional Health, Emotions

## Keywords

Sentiment Analysis, Machine Learning, Naïve Bayes

## 1. INTRODUCTION

We as humans are currently facing great challenges in this era. Aggressive work environments and high stress personal life are taking a toll on our personal health. Social scientists have concluded that people who tend to use social networking aggressively have trouble connecting to real people. When we combine this with added competition and work pressure, we can conclude that humanity is facing historic challenges.

The objective of this project is to improve the emotional health of the user by providing small distractions. These distractions will help the user re-focus on his work/life with a fresher perspective. To do so content will be displayed that will vary from memes (a satirical take on a situation, or scene from popular media), motivational video, funny clips, soothing music, upbeat music and other exercises.

To recognize the emotion from user's input we will use finer sentiment analysis coupled with adaptive media. All of this will essentially help the user to not get bogged down by the pressures of modern day life. Due to increased interference from technology humans are losing the ability to be in real life situations. Various activities will be designed so as to improve emotional health of the user.

## 2. WORKING

Before the user can use the app, significant amount of data about the user needs to be recorded. This data will include name, job type, marital status, and name of spouse (if applicable), names of children (if applicable), favorite television shows and favorite movies. Also needed are names

of two people that the user can turn to in case of emotional need. They will hence forth be mentioned as support people.

## 2.1 Recognizing Emotion

### 2.1.1 Preprocessing

The classifier used is Naïve Bayes classifier based on Supervised Learning. Sentiment analysis is a part of Natural Language Processing (NLP). Our earlier idea was to design the classifier based on a movie review dataset. We realized it doesn't really work for the problem we are trying to solve. Hence we turned to using Twitter data for training our classifier. Twitter being a micro-blogging platform has 140 characters status updates also called as tweets. Thus it is more relevant data where we can find a variety of emotions. The training dataset will be divided in two classes: Positive and Negative, initially. We used a twitter dataset of around 1.5 million tweets where each tweet was labeled either 1 or 0 (where, Positive='1' & Negative='0')

#### 2.1.1.1 Cleaning the Data

This is the first and the most important step when it comes to training. In this era of instant messaging, people have turned towards the use of Internet Slang and short messages. This acts as noise and makes unnecessary data not being identified properly. Removal of emoticons “:)” or “=D” while these seem to help a bit but on testing it didn't really help improve the accuracy.

Tweets contain hash tags, which are denoted like “#word”. It is important to remove these hash tags as those can be highly irrelevant and may not help in training. Hash tags could be anything and is generally used to refer to a particular topic on twitter. Twitter uses “@” for tagging a user in the message. They are represented as “@username”. Such tags should be removed as they serve no purpose and just add to the noise in the data. Twitter messages may contain URL that is a link to different sites on the internet. Thus the removal of links, hash tags and usernames is important so as to decrease the amount of noise or useless words in our feature vector.

#### 2.1.1.2 Bag of Words (BOW)

Every sentence of the dataset will be associated as a feature vector [1]. We maintain lists of all the possible words found in negative tweets and positive tweets. BOW considers every word independent; the word being read now has no connection with the previous word. However, it performs generally very well in text-based classification. Thus a prediction is made based on the probabilities of a word being present in positive and negative list of words. We use the Bag of Words model for further classifying the negative feeling further into: Anger, Depressed- Professional, Depressed-Personal and Anxiety.

We could not find any data set which could have the tweets labeled into the above mentioned categories. Hence we take a

bag of words approach wherein we use a predetermined set of words for each class.

### 2.1.1.3 N-grams

They are essentially ordered set of words with length  $n$ . Ngrams perform well for locating syntactic patterns, especially the ones we need i.e. negations e.g., “not happy”. [1]. Negation is an important component for the analysis of emotion in text as slight changes can make drastic differences. For example, the sentence “I’m not sad” should be classified into the positive category and not into negative. Words like “very” or “definitely” don’t provide a lot of sentiment information on their own, but phrases like “very bad” or “definitely recommended” increase the chances of a tweet being negatively or positively biased. By including bigrams we are able to capture this information about adjectives and adverbs [3]. In our implementation we take into consideration only bigrams. We extracted bigrams from 100,000 tweets.

### 2.1.1.4 Part of Speech Tagging

Part of speech tagger is used to determine the part of speech of each word in a tweet. The part of speech of a word will vary depending on its usage in the sentence. We delete part of speech like Interjections, Cardinality, Verbs, Determiners and Pronouns [3].

This help remove words are not necessary, the ones that do not fit in as a feature that can be used for training. For example, take a sentence like “It was a hard day at work”. This will be POS tagged as “it/PRP was/VBD a/DT hard/JJ day/NN at/IN work/NN” (where, PRP- Pronoun, DT- Determiner, NN-Noun, IN- Interjection, VBD- Verb). Out of this the useful words extracted are (hard, day, work) which contribute to our feature vector of tweets.

### 2.1.1.5 Negation Handling

Negation handling was among the factors that contributed significantly to the accuracy of classifier that we made [2]. A major issue faced during the task of sentiment classification is that of handling negations. Since we are using each word as feature, the word “good” in the phrase “not good” will wrongly be contributing to positive sentiment rather than negative sentiment as the presence of “not” before it is not taken into account. We extract this “not good” from the sentence with the help of Bi-grams. To solve this problem a simple algorithm for handling negations using state variables. The algorithm uses a state variable to store the negation state. It transforms a word followed by a not or n’t into “!” concatenated with the word. Thus when something like “not good” is detected it will be transformed in the “!good” and added to the corresponding feature vector. Thus to classify “!good” as negative, we need a huge dataset where there is a possible occurrence of this combination. Thus it is suggested to use Bigrams only when a good amount of data can be used for training.

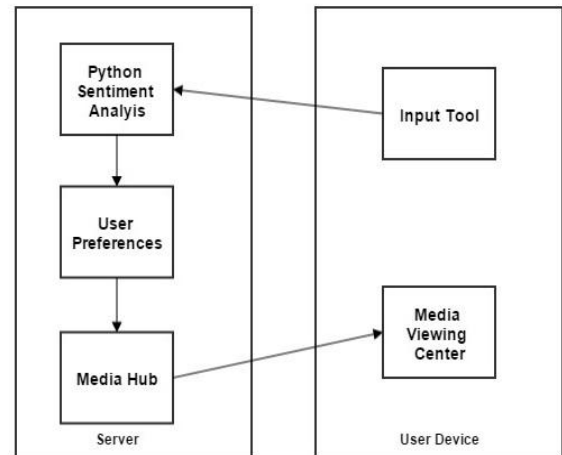
## 2.2 Training Results

We trained our classifier using variable size of training data and found some interesting results. In the below table we show our the test results using different number of tweets used for training the classifier and what accuracy we managed to achieve considering only unigrams and unigrams + bigrams. We use 20000 tweets for testing the accuracy.

**Table 1. Training Results**

No. of tweets for Training	Accuracy (%)	
	Unigram	Unigram + Bigram
50,000	67.32	69.39
100,000	77.12	78.24

Using the results of the above table it is clear that there is the accuracy improves as the size of training set increases. Also, there is a slight improvement when we use bigrams and unigrams.



**Fig 1: Flow of information**

## 2.3 Profile Management Part

Once the exact emotion is found the next part is to display a specific media to counter the emotion. The negative emotions are primarily divided into four categories namely, Anger, Depressed- Professional, Depressed- Personal and Anxious. The media displayed for the each case is different. Unlike other emotions, positive-joyful state is usually desired; so no media will be displayed to counter it.

### 2.3.1 Joy

Joy can be defined as a state of happiness which lasts for a short period of time. When this state is detected the user will have to capture a selfie (image of the user captured by the user).

This selfie will later be used when there user is depressed state. The objective of this exercise is to remind the user that even if he/she is sad right now, some time ago he/she was joyful.

The objective is to put the user in a better state of mind by using personal experience.

### 2.3.2 Anger

After this state is detected the user can be given following responses

- Funny meme based on the user's personal interests.
- Short funny clip
- 1 minute symphony track
- Psychologically proved exercises

### 2.3.3 Depressed – Professional

This type of emotion will be identified by tracking keywords such as "goal, target, boss, work, failure". The output can be

- Motivational images
- Motivational success stories
- Upbeat music
- Activities like calling up support people, listing down causes of failure

### 2.3.4 Depressed - Personal

This type of emotion is linked to personal reasons. The output can be

- Joy selfies
- Activities like calling up support people.
- Funny memes based on personal interest.
- Funny clips

### 2.3.5 Anxiety

This type of emotion is for situations when the user is tensed due to pressure-crunch situations. The output possible is

- Soothing music
- Simple breathing exercises
- Funny clips
- Activities like calling up support people

## 2.4 Intelligent Results

Let's consider that the user is 'Depressed – Professional' every day between 11 am – 12 noon. The user's depression is identified and specific media is shown. But the user still requests for more media. That means that despite showing media the user's emotional state does not show change. Hence the application needs to use more powerful measures such as motivational video or initiate activities like calling support people.

The application needs to be intelligent so as to understand the hours in which the user is most depressed. For such requirements a 1-Dimensional matrix needs to be created. It will represent the hours of the day. Such a matrix will be specific for each emotion. All the values in the matrix are assigned the value zero. The objective here is to increment by 1 whenever the user requests media in the hour. Hence over time the value will increase more in certain parts of the day, than in other parts of the day i.e. value for a certain hour will be higher than the value of other hours.

There will be threshold variable assigned to the user. There will also be a function that periodically will scan the values in the 1D dimensional matrix and compare them to the threshold variable. If the value is more than the threshold variable then the user will need more powerful media henceforth. The threshold variable will then be changed to a higher value.

Hence the media will change as user's requirement changes.

## 2.5 Other Features

Other Features are designed in conjunction with the research that "happiness is being thankful for existing things". Hence at the end of each day the user will be asked "What is he/she thankful for". A question like this causes the user to assess his/her life and try to note the positive aspects of it.

The user will be also be given weekly challenges that he/she will have to complete within a week. The objective of these challenges is to expose the user to new things that will break the existing emotional barriers. The goal is to strengthen existing emotional ties and make new ones. The difficulty of tasks will increase every week.



Fig 2: Sample results

## 3. RESULTS

In the D.J. Sanghvi Computer Department Project Display 2015, we showed the application on an android phone, to 35 people. The people here include professors and students. There were more students than professors.

We asked them a series of questions as a survey with regard to the application they just used. Here is the result of that survey.

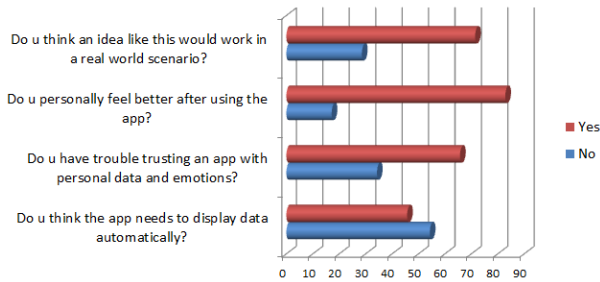


Fig 3: Results of the Survey, Part 1

On whom should the app focus more?

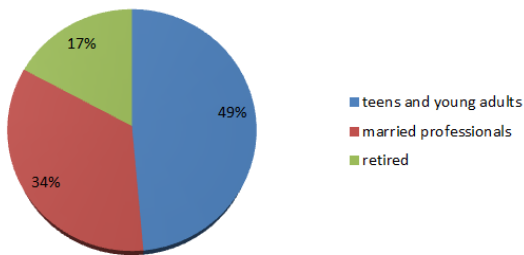


Fig 4: Results of the Survey, Part 2

#### 4. CHALLENGES

The app relies a lot on user's information. If the user is uncomfortable with that, then the process cannot function.

People ideally would like the application to show results automatically without typing. But for such level of automation we either need portable heart monitors or portable brain wave detectors. Both of these devices are not cheap hence have to be bought separately. Also wearing these devices all the time is not possible.

Getting new content every day is challenging. So in an ideal scenario if the app can tie up with applications like 9gag and YouTube, then user will never be short of data.

#### 5. FUTURE SCOPE

As seen in Fig 3, participants want the app to display data automatically whenever the user is emotionally unhinged. One way to do so is to use a portable heart monitor or pulse monitor. As user's emotions change the heart beating patterns change as well. Such portable heart monitors are part of smart watches released nowadays. So heart monitors can detect certain patterns and trigger the app. The app will download content, and show a notification as the media download is done.

One great upside to this is that the user won't have to type anything or wait for content to download. The downside currently is that smart watches are expensive; as of now so mass acceptance will be less.

As time passes and more data is collected about the user, the application's performance will improve over time. Today's smartphones are available with smart AI based assistants. Some examples are Cortana, Siri and Eva. The application can

be integrated with these smart assistants that make the experience more seamless and realistic.

Usually Smart Assistants already have most of the personal information about the user which will help keep track of a user's mental state and also help in learning more things about the user, thus making the use of the application more efficient.

#### 6. ACKNOWLEDGMENTS

We would like to thank our honorable principal Dr. Hari Vasudevan of D. J. Sanghvi College of Engineering and Head of Department of Computer Engineering, Dr. Narendra Shekokar for giving us the facilities and providing us with a favorable environment for working in college. We would also like to thank our friend Pratik Mehta for his help in the project. Our project guide Prof. Sindhu Nair has provided us great help and insight.

#### 7. REFERENCES

- [1] Soumaya Chaffar and Diana Inkpen, "Using a Heterogeneous dataset for Emotional Analysis In Text," Advances in Artificial Intelligence, Lecture Notes in Computer Science Volume 6657, 2011
- [2] Vivek Narayanan, Ishan Arora and Arjun Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model," Intelligent Data Engineering and Automated Learning – IDEAL 2013, Lecture Notes in Computer Science Volume 8206, 2013
- [3] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," Proceedings of EMNLP, 2002
- [4] Andrew B. Armstrong, "Basis for Ratio Advice", "Altering positive/negative interaction ratios in relationships of mothers and young children", Utah, Proquest, Umi Dissertation Publishing
- [5] Carlo Strapparava and Rada Mihalcea, "Annotating and Identifying Emotions in Text", ACM SAC, 2008
- [6] Alena Neviarouskaya, Helmut Prendinger and Mitsuru Ishizuka, "Recognition of Fine-Grained Emotions from Text: An Approach Based on the Compositionality Principle", Modeling Machine Emotions for Realizing Intelligence, 2009
- [7] Martin D. Sykora, Thomas W. Jackson, Ann O'Brien and Suzanne Elayan, "Emotive Ontology: Extracting Fine-Grained Emotions From Terse, Informal Messages", IADIS International Journal on Computer Science and Information Systems, 2013, Vol. 8, No. 2
- [8] Saima Aman and Stan Szpakowicz, "Identifying Expressions of Emotion in Text", Text, Speech and Dialogue, 2007
- [9] Svetlana Kiritchenko, Xiaodan Zhu Xiaodan and Saif M. Mohammad, "Sentiment Analysis of Short Informal Texts", 2014, Journal of Artificial Intelligence Research", Vol.50