

An approach to predict emotional state using printed document

Mohit Kumavat¹, Amrita Kumbhar², S.V. Kedar³, Suraj Menon⁴, Vineet Prasad⁵, Omkar Parit^{6*}

^{123456*} Department of Computer Engineering, RSCOE, S. P. Pune University, Pune, India

*Corresponding Author: kumavatmohit007@gmail.com, Tel.: +91 9762938996

Available online at: www.ijcseonline.org

Accepted: 22/Jun/2018, Published: 30/Jun/2018

Abstract— An emotion is a strong feeling deriving from one's mood, circumstance, or relationships with others. They include the perception, response and interpretation of the feelings related to the experience of any specific situation. Emotions are the ones which connect the feelings, actions and thoughts. When we interact with other people, it is important to give cues to help them understand how we are feeling. Text analysis is one of the techniques, which helps us to understand a subject in a proper way through keywords used by him/her. By studying the keywords, we can paint a picture of the writer's emotional fears, honesty, outlays, mental state and many other personality characteristics. In this paper we aim to analyze the keywords to determine the person's emotion levels by using the keyword based method for emotion recognition. This will help to identify people with stress and depression and the ones who need counseling to come out of such emotions.

Keywords— Emotions , Emotion recognition, Text, Keywords, Optical Character Recognition.

I. INTRODUCTION

Emotion is any conscious experience characterized by intense mental activity and a certain degree of pleasure or displeasure [1]. Emotions have a large impact on a person, his behaviour and attitude. The emotions are related with personality, mood, outlook, character, and inspiration. Emotion recognition can help to uplift one's emotional state. Study has been done on various approaches to recognize current emotional state of a person. Facial expression, speech recognition, handwriting sample, body pose, keyword detection are some of the approaches towards emotion recognition.

In this paper we will study approach using keyword detection. As most of the data these days are available in the form of online blogs, posts. People use various social platforms to express their feelings and updates. This vast information can be easily accessible and be utilized to find out emotional state using keyword detection method. This paper can be implemented in a team by analyzing emotion and identifying highly profitable group of people so as to get work done fast and in a better way.

In paper [4] authors have surveyed the different methods for emotion detection using keywords. They have tried to improve the limitation capabilities used for text based emotion detection. Identifying newly evolved vocabulary set, systematic emotion ontology which is based on the OCC model and a collective method to predict various emotions.

In paper [3] authors have used Keyword Based Approach to recognize emotion from text. The author has worked on the following emotions: love, joy, sadness, anger, fear and surprise. The Proposed Emotion Detector System produces output with the overall accuracy of 79.57%. Individual

accuracy achieved for love is 75%, for joy is 70.58%, for sadness is 80.95%, for anger is 69.69%, for fear is 81.25%, for surprise is 100%.

In paper [5] authors have used SVM and NB classifier for recognizing emotion from text document. Here emotion is recognized particularly for Punjabi script. The dataset considered is standard 'HC Corpora' dataset for Punjabi language.

The section I. describes about what is emotion and a brief detail about some works in emotion detection using keyword based methods. Section II. covers different methods of emotion detection through text viz. Keyword method, Lexical affinity based method, Hybrid method and Learning based method. Section III. covers the methodology used which is the keyword based method. It uses the Naïve Bayes classifier algorithm. It also gives a detailed architecture of the proposed system and its working. Section IV. covers the experimental results in which we have acquired an overall accuracy of 70.37%. Section V. gives the conclusion and overall summary of the working system.

II. RELATED WORK

There are various approaches for emotion detection. Emotion detection through facial expression, text, speech, body pose are a few. Each of these approaches have some pros and cons. Study shows that Emotion detection using text is more accurate with minimum error rate.

Also increase in number of social media, blogs sites and some other platforms encourages people to express their feelings through text. This has increased scope of emotion

detection using text recently. Text based methods can be done using methods like keyword detection, Learning-based method, Lexical affinity method and hybrid method [3].

A. Keyword method :

This method works on the basis of predefined keywords. A category of keywords is formed to detect emotions such as disgust, happiness, sadness, anger, fear, surprise etc. Based on occurrences of these keywords an emotion class of the text document is found.

B. Lexical affinity based method:

Lexical affinity is an approach in which a probabilistic 'affinity' is assigned for individual emotion to arbitrary words by picking up keywords which detect emotions. These probabilities belong to a linguistic corpora. However they have some disadvantages; e.g. keyword 'accident' generally indicates a negative emotion, but it would incorrect result for phrases like 'I met my girlfriend by accident' or 'I avoided an accident' which have a totally different meaning.

C. Learning based method:

Learning-based methods make use of previously trained classifier to detect emotions, which uses machine learning techniques to determine which emotion class the input text belongs to.

D. Hybrid method:

Here different methods are combined to get best of different methods for more accuracy.

In this paper we are using Keyword based method. This is the simplest and straight forward way for emotion detection using text. Though this method is simple, if the text itself is very ambiguous it becomes difficult to detect the exact meaning of text. Keyword-based emotion detection methods have the following limitations ^{[4][5]}.

1) Ambigutiy in keyword detection:

The keywords method of emotion is generally quite simple, however the meaning of the keyword can be ambiguous and vague.

2) Incapabiity of recognizing sentences without keywords:

Keyword-based method is totally dependent on the set of keywords. Therefore, sentences without any keywords would indicate that they are emotionless, which is obviously wrong.

III. METHODOLOGY

In this approach a database will be maintained having key-value pair of various keywords and its corresponding emotion. A database of six basic emotions such as Happy, Sadness, Anger, Fear, Disgust and Surprise is created here. Here input that will be taken is in image format. Optical Character Recognition (OCR) is an image to text conversion technique. The printed text is recognized by a computer using OCR. First the text is scanned character-by-character, analysis of the scanned-in image is done next, and then the character image is translated into character codes, such as ASCII, which is commonly used data processing [2].

A. Algorithm

Naive bayes classifier :

The Bayesian Classification comes under the supervised learning method. It makes use of statistical classification method. It considers an inter-defined probabilistic model that allows us to calculate uncertainty in the model by determining probabilities of the outcomes. It can be used to solve predictive problems as well as diagnostic problems.

The Bayes Theorem:

$$P(h/D) = P(D/h) * P(h) / P(D)$$

Where,

P(h) : Prior probability of hypothesis h

P(D) : Prior probability of training data D

P(h/D) : Probability of h given D

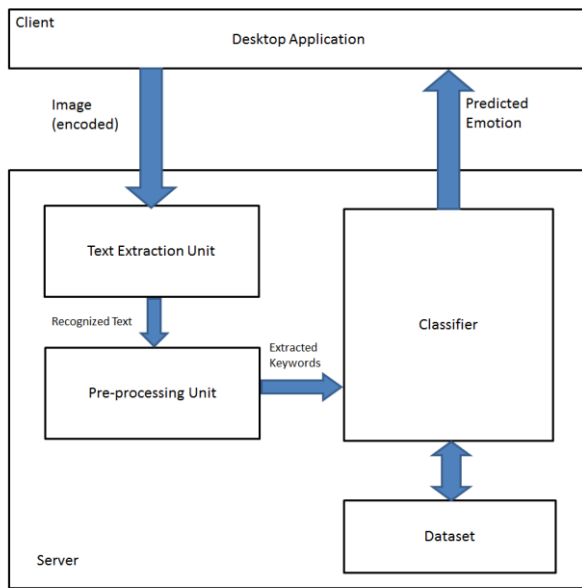
P(D/h) : Probability of D given h.

B. Proposed system

The proposed system consists of Architectural diagram, Algorithm and working of entire system. It gives detailed working of system from taking input then pre-processing of input, fetching keywords out of it and then generating output using Naïve Bayes Classifier. Proposed system gives basic idea of how the system will look like.

1) Architecture

Figure 1. Architecture Diagram



2) Algorithm:

Case 1) If the input is an image:

1. Capture an image of the printed text sample and store it in the SD card.
2. Mount that SD card on raspberry-pi hardware.
3. Browse the image through interface provided by the software.
4. Click on “Fetch OCR and Sentiment Result” button.
5. The request is sent to the server in json (key-value pair) format.
6. Apply the following steps (6.1 to 6.3) for every image of testing dataset.
 - 6.1. Encode the image into Base-64 format and this response is sent in json format to the server in the back-end.
 - 6.2. Server then decodes the image and sends it to OCR (Tesseract software) to extract the text.
 - 6.3. The extracted text is then pre-processed by removing Stemming and Stop words.
7. The Naïve Bayes’ training dataset is prepared as the steps described below:
 - 7.1. Collect samples of every emotion in form of words or sentences.
 - 7.2. Create .csv files for every emotion defined.
 - 7.3. Segregate each sample and store it in appropriate .csv file.
8. The Naïve Bayes’ classifier compares each input sample with the training dataset samples using Naïve Baye’s approach.
9. The most appropriate emotion is detected and displayed.

Case 2) If the input is a typed message:

1. The user types the message in a form created on webpage.

2. The message is then sent in the form of a request to the controller in back-end.
3. The controller then extracts the text in the message and sends it for pre-processing.
4. The steps from 6.3 in the above algorithm are then followed and the result is displayed.

3) Working:

Here we can provide input as a text or an image. In case of image input will be encoded into Base64 for transmission from client machine to server (Raspberry pi). At server side encoded input will be decoded to get original image. Optical Character Recognition is implemented using Tesseract. Tesseract will fetch text from image. Then preprocessing will be done on the extracted text to remove stop words and stemming’s. This will give us set of keywords from the input. Naïve Bayes’ Classifier will classify the keywords based on key-value pair stored in database. Based on the probability distribution it will find out emotion on the input text.

Keywords which are detected but do not exist in database will be mapped to class of detected emotion in database.

Font OcrB is used for better accuracy.

IV. RESULTS AND DISCUSSION

In total we have tested 54 samples, out of which 38 gave the correct result. For happy 12 samples were tested and 10 were correct. For Sad 8 samples were tested out of which 6 were correct. For Angry 10 samples were tested out of which 8 were correct For Disgust 6 samples were tested out of which 2 were correct. For Fear 10 samples were tested out of which 7 were correct. For Surprise 8 samples were tested out of which 5 were correct. We have gained an accuracy of 70.37% by testing the above samples.

The conclusion matrix is as follows:

Table 1. Conclusion Matrix

Expected Result	Obtained Result						Total
	Happy	Sad	Angry	Disgust	Fear	Surprise	
Happy	10	1	1	0	0	0	12
Sad	0	6	1	0	0	1	8
Angry	0	0	8	0	1	1	10
Disgust	0	0	3	2	0	1	6
Fear	0	1	0	0	7	2	10
Surprise	3	0	0	0	0	5	8

$$\frac{10+6+8+2+7+5}{12+8+10+6+10+8} = \frac{38}{54} = 70.37$$

V. CONCLUSION

In this project we have presented methods used for emotion recognition using printed documents. The detection of emotions will be done by using keywords. Text analysis is one of the techniques, which helps us to understand a person in a better way through keywords used by him/her. By studying the keywords, we can paint a picture of the writer's emotional fears, outlays, honesty, mental state and many more personality characteristics. There will be counseling provided if any negative emotions are encountered. This project can be used in offices and workplaces to improve the psychological state of the employees. This can be used in prevention of suicides among student under pressure or any other professional in general. A training dataset will be provided beforehand in which a data dictionary will be provided. Keywords will be present in the dictionary. Keywords will be extracted from the newly acquired samples and will be compared with keywords in the dictionary and hence the emotion will be recognized.

ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on 'An approach to predict emotional state using printed document'. We would like to take this opportunity to thank our internal guide and Head of Computer Engineering Department, JSPM's Rajarshi Shahu College of Engineering, Tathawade Prof. Seema Kedar for giving us all the help and guidance we needed. We are really grateful to her for her kind support. Her valuable suggestions were very helpful. In the end our special thanks for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project.

REFERENCES

- [1] Prof. S.V. Kedar, Dr. D. S. Bormane, "Automatic Emotion Recognition: A Systematic Review", IETE, India, 2016.
- [2] Shiv Naresh Shivhare, Shakun Garg, Anitesh Mishra, "Emotion Finder: Detecting Emotion From Blogs and Textual Documents", International Conference on Computing, Communication and Automation, India, 2015.
- [3] Edward Chao-Chun Kao, Ting-Hao Yang, Chang-Tai Hsieh, Von-Wun Soo, "Towards Text-based Emotion Detection A Survey and Possible Improvements", International Conference on Information Management and Engineering, Taiwan, 2009.
- [4] C.-H. Wu, Z.-J. Chuang and Y.-C. Lin, "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), vol. 5, issue 2, pp. 165-183, 2006.
- [5] Sheeba Grover, Dr. Amandeep Verma, "Design for Emotion Detection of Punjabi Text using Hybrid Approach", India
- [6] D. Sanchez, M.J. Mart'in-Bautista, I. Blanco, "Text Knowledge Mining: An Alternative to Text Data Mining", in IEEE International Conference on Data Mining Workshops, 2008.

- [7] Shiv Naresh Shivhare, Saritha Khethawat, "Emotion Detection from Text", Second International Conference on Computer Science, Engineering and Applications (CCSEA-2012), India, 2012, ISBN: 978-1-921987-03-8.
- [8] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, "Recognition of Emotional States in Natural human-computer interaction," in IEEE Signal Processing Magazine, vol. 18(1), 2009.

Authors Profile

Mohit Kumavat pursued Bachelor of Engineering in Computer Science from Savitribai Phule Pune University, Pune in 2018. He has published 2 Patents in Indian Patent Journal in 2017 and 2018.



Amrita Kumbhar pursued Bachelor of Engineering in Computer Science from Savitribai Phule Pune University, Pune in 2018. She is currently working as Associate Consultant at Syntel, Pune.



Seema Kedar pursuing Ph.D from Savitribai Phule Pune University. Associate professor and head of computer department in JSPM's RSCOE, Pune.



Suraj Menon pursued Bachelor of Engineering in Computer Science from Savitribai Phule Pune University, Pune in 2018.



Vineet Prasad pursued Bachelor of Engineering in Computer Science from Savitribai Phule Pune University, Pune in 2018.



Omkar Parit pursued Bachelor of Engineering in Computer Science from Savitribai Phule Pune University, Pune in 2018.

