iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** I **Month of publication:** January 2020

DOI: http://doi.org/10.22214/ijraset.2020.1042

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 8 Issue I, Jan 2020- Available at www.ijraset.com

Text and Voice Input to Emotional Analysis for AI Clients

Simran Bhake[1], Shruthi Borkar[2], Siddhesh Swami[3], Pawan Yadav[4]
[1, 2, 3, 4]*Computer Department, RAIT*

Abstract: ***In the past few decades, the influence of technology and especially the internet has grown rapidly. People are interacting with one another via machine and also the ways Human-Machine interaction is undergoing rapid changes and reforms. Textual and voice data are main tool of Human-Machine interaction. In order to make communication and interaction with machine as real as possible, emotional status of humans must be recognized by Machine. Due to this, there is heavy research going on in the field of emotional analysis. Humans can express their emotion in various ways but it is mainly done through text and voice/audio over the internet. Emotions are extracted from various data entered by user on blogs and social sites and emotion estimation is done. This research paper is focused on emotion detection techniques based on inputs extracted from blog and social sites and their relative merits.***

I. INTRODUCTION

In human interaction and communication, emotions play a very crucial role. It has a large part in our ability to communicate effectively.Similarly the more effectively the machines are able to recognize human emotions, the more improved services it can offer. Detecting Emotional state of a person can be challenging but also very crucial. Emotions can be detected through various ways from human i.e facial expressions, gestures, speech and written text. Emotion is expressed as joy, anger, surprise, hate, fear and so on. In this paper we discuss the text and voice based input in emotional analysis.

The detection of emotion from sentence is the most important part of emotion analysis. Emotion detection from text consists of extraction of emotional class method. The emotion class are those words which distinctly denotes the emotional expressions of the sentence to a particular emotion. There are various methods to detect the emotions from text. We will discuss a few of them further.

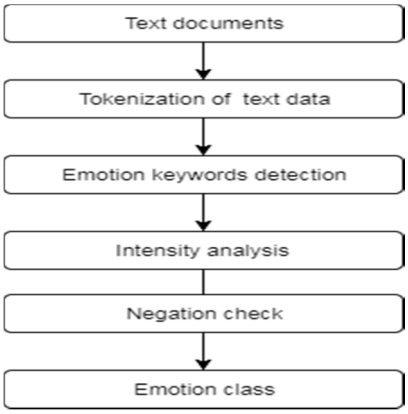
Detection of emotions from speech consist of analysing the speech signal by the user and using emotional class method. Emotions have also direct influence to the nervous system, hence the heart rate is also influenced by them. Speech signal are also a representative of heart rate of the user as the heart rate is affected by speech. Due to this, the speech signal can be used to detect the emotion more accurately than textual data. Currently emotional recognition is used in various fields like in medicine, e-learning, Monitoring, Entertainment, Law, Marketing, etc. The better emotion detection system will be more the benefits of it will be provided respective fields. Hence a lot of research going on to improve the detection of emotions from humans.

II. TEXT BASED EMOTION RECOGNITION METHODS

The different text based emotion recognition techniques are: Keyword spotting method, Lexical Affinity method(lexicon based), Learning based detection(SVM based) and Hybrid based detection.

A. Keyword Spotting Method

In this method, classification of emotions is done by searching for the specific emotional keywords in the input sentence.This method finds the occurrences of keywords from a given data set.These words are then classified into different categories such as sadness, boredom, happy, anger, disgust, fear, etc



The input is a text document and the output is generated as an emotion class. Keyword spotting technique consists of five steps. In the first step, input data from text is converted into tokens. From these tokens emotion words are identified. Afterwards analysis of the intensity of emotion words is performed. The sentence is checked whether negation is present in it or not. Then finally an emotion class will be displayed as the required output. However, classification methods based on only keywords suffer from (1) the enigma in the keyword definitions in the context that a word can have different meanings according to usage. (2) the incapability of recognizing emotions within sentences that do not contain emotional keywords (3) the lack of linguistic information.

B. Lexical Affinity Method

Keyword spotting method is a straightforward method which detects specific emotional keywords. The lexical affinity is an extension of keyword spotting method. In this approach, instead of detecting predefined emotional keywords from text, a probabilistic affinity is assigned for a specific emotion to arbitrary words. These probabilities are often a part of linguistic corpora.The main disadvantage is that this method misses out emotional content that reside deeper than the word level on which this technique operates.

For example, “The book lacks a good story.”

This method spots the keyword ‘good’ and identifies the sentence as positive emotion.But in reality,the sentence is of negative emotion.

C. Learning Based Method

Learning-based techniques attempt to perceive feelings in light of a past prepared classifier/comes about, which mapped with different machine learning classifiers, for example, support vector machines , specific statistical learning methods and decision trees, to identify which feeling classification/class should the information content has a place. This approach confront challenges like these strategies may order sentences into just two classifications as a result of lacking highlights other than feeling catchphrases, which are negative and positive. Waste make utilize the possibility that feelings are identified with human mental states which are caused by some enthusiastic occasions. This implies the human personality begins with starting mental state and moves to another state upon the event of a specific occasion. This thought is actualized utilizing Concealed Markov Show where each sentence comprises of many sub-thoughts and every thought is dealt with an occasion that makes a progress a specific state. The grouping of occasions in the sentence is trailed by the framework also, decides the feeling of the content. The framework accomplished F-score of 35% when tried on the ISEAR (International Survey on Emotion Antecedents and Reactions)) dataset, where the best exactness accomplished was 47%.

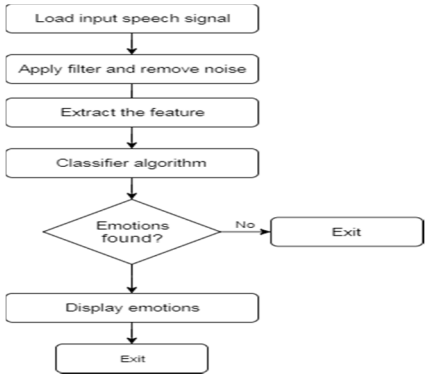
D. Hybrid Based Detection

In Hybrid method, emotions are detected by utilizing the combination of Keyword spotting method and learning patterns collected from training datasets.This method results in higher accuracy than both the methods used individually.

The most symbolic hybrid system so far is the work of Wu,Chuang and Lin,that utilizes a rule-based approach to extract semantics related to certain emotions and Chinese lexicon knowledge to extract attributes.These semantics and attributes are related with emotions as emotion association rules.Subsequently,these rules replace original emotion keywords and serve as a training dataset for their learning module.

III. VOICE BASED EMOTION RECOGNITION

The Steps involved in Emotion Detection through Voices are listed in the following flow chart.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 8 Issue I, Jan 2020- Available at www.ijraset.com

A. Voice Input

For any model to work precisely, the input should be taken with great care. The efficiency of the system is directly proportional to the quality of the database. The data may contain different types of input of different genders and ages and can contain collection of acted speech and real world data.

B. Feature Extraction

The most Important part in emotion detection is the extraction of features from the voice input. After collection of input database features are extracted from the database. The commonly used features are pitch, energy, MFCC, LPCC, formant. Out of them MFCC is the most widely used in emotion recognition.

C. Classifier Methods/ Algorithm

By the extracted feature the classifier to classify the signal based on various properties. There are various methods. Some of them are listed Below.

1) *Neural Network Based Approach:* The Neural Network is a computer system modelled on the human brain and nervous system. Human brain consists of large number of neurons linked with each other in a highly complex and nonlinear manner. It perform various complex task which today's supercomputers cannot perform.

Steps for Emotion Recognition process are explained below.

a) *Step 1:* The extracted feature is and input data is trained with the help of neural network.
b) *Step 2:* In this step the Fuzzy theory approach is applied to check the results from the simulated data. Create The Network Input¬P Target¬T
c) *Step 3:* The network simulation is performed with the help of neural network.
d) *Step 4:* If Output Is Equal To Target then detect the emotion and display the emotion and also play the voice.
e) *Step 5:* If output is not equal to the target then it will end.

The main advantage is Neural network are capable of solving much more complicated recognition task. They can handle low quality, noise data and speaker independent data.

They can achieve greater accuracy in detecting hot, anger, neutral and happy and sad emotions.

The drawback is that they do not scale well in large vocabulary system and their vocabulary is limited.

Their performance is highly depends on the emotional speech samples.

2) *K-Means Clustering:* The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity. Here Based on features, the speaker's emotion is distinguish.

The objective function can be represented by

$$J=\sum_{i=1}^{k}\sum_{j=1}^{n}(\|S_n^{(i)} - C_i\|)^2$$

Here, J is the Number of emotions containing all states of emotions. C_i is the centre of cluster obtained by partitioning features in k cluster. $\|\ \|$ is norm representing the distance between C_i and the data point $S_n^{(i)}$.

Steps involved in K-means algorithm are

a) From each **k** feature points, select the centroid.
b) Obtain **k** cluster by iteratively repeating the procedure. In the process, allot all the source data point to the respective nearest centroid.
c) The centroids are updated by estimating the cluster centers iteratively, until further variation in cluster center is manifested.

The benefits in this method is that if the input variables are huge then k-means clustering computes faster if k is kept small. The main challenge in K-Means is that it cannot pick up on all the information contained in the feature sets, unless we add some bias to the features.

3) *Support Vector Machines (SVMs):* SVM is a very simple and efficient classifying algorithm which is used for pattern recognition (Fig. 5) (Pao et al., 2006). Support Vector Machines were introduced by Vladimir Vapnik in 1995.The main idea is to transform the original input set into a high dimensional feature space by using a kernel function and then, to achieve optimum classification in this new feature space, where a clear separation among features obtained by the optimal placement of a separation hyperplane under the precondition of linear reparability

Steps

a) Labelling the extracted feature value in different classes and storing in database with corresponding labels.

b) Feature values of speech signals that are obtained from emotional databases speech utterances are fed into SVM along with their class labels.

c) Developing SVM model for each emotional state using feature values.

d) Taking real-time input and comparing with predicted values.

Benefits are that it can produce accurate and robust classification results on a sound theoretical basis, even when the input data are non-monotone and non-linearly separable.

The accuracy of results does not rely on the quality of human expertise judgement for the optimal choice of the linearization function of non-linear input data.

The main obstacle is for very high precision, you need a lot of memory since you have to store all the support vectors in memory and this number grows sort of linearly with the training dataset size.

4) Hidden Markov Model (HMM): A hidden Markov model (HMM) allows us to talk about both observed events (like words that we see in the input) and hidden events (like part-of-speech tags) that we think of as causal factors in our probabilistic model. In this approach, variations in speech are modeled statistically (e.g., HMM), using automatic learning procedures. This approach represents the current state of the art. Modern general-purpose speech recognition systems are based on statistical acoustic and language models. According to Deller et al. (1993), the states in the HMM frequently represent identifiable acoustic phonemes in speech recognition. The number of states is often chosen to roughly correspond to the expected number of phonemes in the utterances. However, the optimal number of states is best determined through experiments as the relationship of the number of states to the performance of the HMM is very imprecise. The Steps involved are learning from the various input. Decoding the input and evaluation of the Emotions. The main gains are the word sequence and a pronunciation dictionary and the HMM learning process can automatically determine word and phone boundary information during training. This means that it is quite straightforward to use large training corpora. The demerits is Hidden Markov Model is expensive, both in terms of memory and compute time.

IV. ANALYSIS

From the Neural Network approach, the Deep-Learning (DL) algorithm has recently shown great promise in many benchmark image- and speech-recognition challenges.

The important limitation of DL is possible overfitting and lack of generalisability. When faced with limited data, K-means based approach is a powerful alternative to deep-learning algorithms. K-means clustering is much simpler than Neural Network approach. The most common application of HMM is in speech recognition. One of the main advantages of HMMs is their ability to model nonstationary signals or events. The use of Mel-frequency cepstral coefficients (MFCC) on SVM classifier, not only gives the best overall performance on 8 kHz sampling frequency, but also shows consistent performance for all the emotion classes, compared to other classifiers and feature combinations with less computational complexity. HMM speech recognition can be enhanced using SVM and Neural classifiers SVM is used to binarize the features of the input sample and Neural to finally classify the entire architecture.

V. CONCLUSION

This paper presented an overview and analysis of emotion recognition techniques by using text and voice inputs. We found that the efficiency of algorithms changes with the requirements. K Means algorithm is best in the applications where speed is an important factor. SVM is best in the applications where space(data storage) is not a constraint. HMM provides best result, but expensive both in terms of memory and computation time.

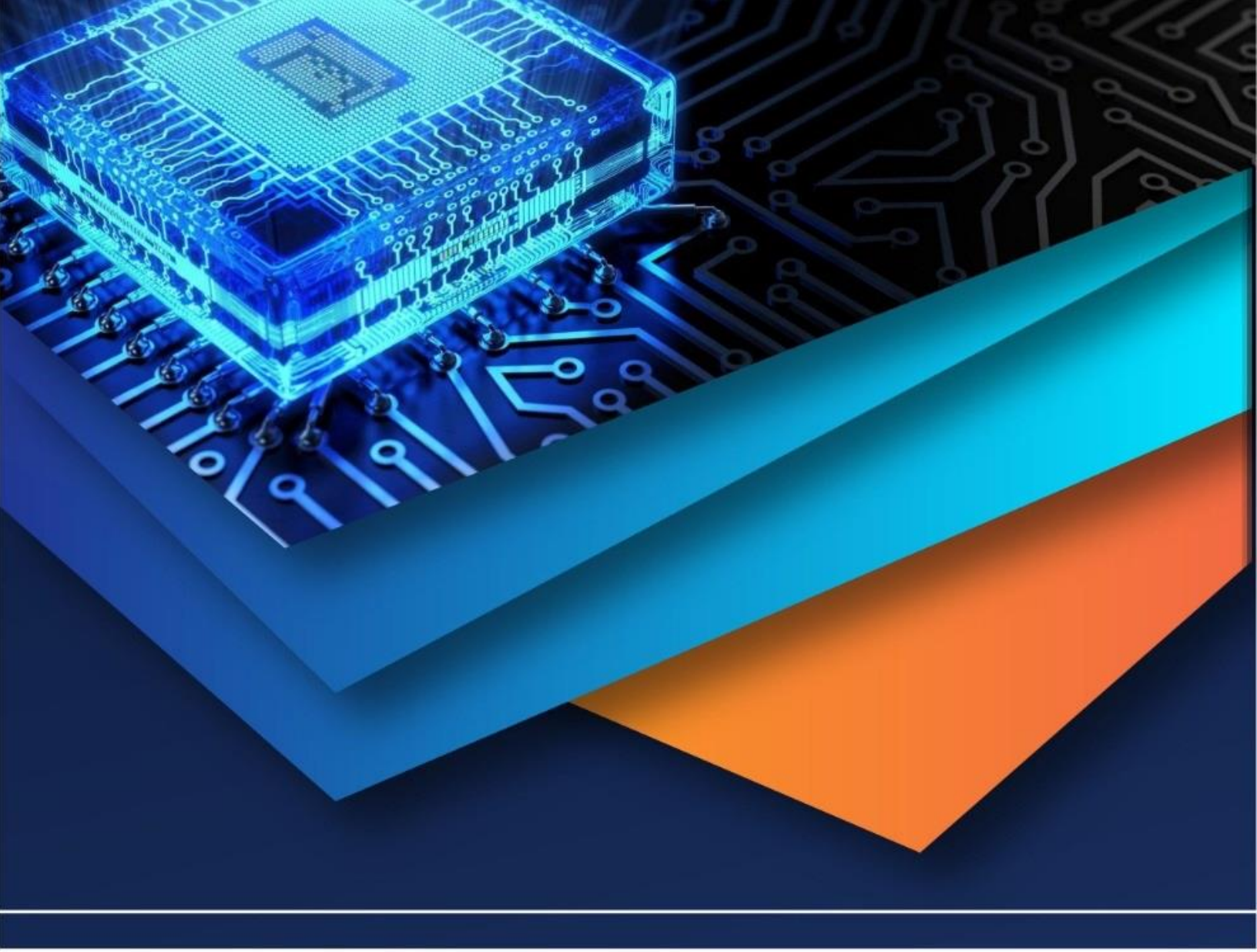
Current research going across the globe will help us to get an idea about the largeness of the field.

VI. ACKNOWLEDGEMENT

We would like to thank CSI-RAIT Committee, Computer Engineering Department and CSI research team mentors for their constant support.

REFERENCES

[1] Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, Von-Wun Soo. "Towards Text-based Emotional Detection:A survey and possible Improvements,"in International Conference on Information Management and Engineering,2009.
[2] 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, Jun. 2006, pp. 165-183, doi:10.1145/1165255.1165259.
[3] Arti Rawat, Pawan Kumar Mishra," Emotion Recognition through Speech Using Neural Network" International Journal of Advanced Research in Computer Science and Software Engineering , Volume 5, Issue 5, May 2015
[4] Hemanta Kumar Palo, Mihir Narayan Mohanty, Mahesh Chandra, "Emotion Analysis from Speech of Different Age Groups" , Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering pp. 283–287,ACSIS, Vol. 10 ISSN 2300-5963
[5] J. Kaur, and S. Vashish, "Analysis of different clustering techniques for detecting human emotions variation through data mining," International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), vol. 3, iss. 2, pp. 27– 36, Jun. 2013
[6] Inderjeet Kaur, Dr. Rakesh Kumar , Palwinder Kaur," Speech Emotion Detection Based On Optimistic – DNN (Deep Neural Network) Approach"
[7] Miss. Aparna P. Wanare," Int. Journal of Engineering Research and Applications", ISSN : 2248-9622, Vol. 4, Issue 7(Version 4), July 2014, pp.74-78
[8] Laura Auria , Rouslan A. Moro," Support Vector Machines (SVM) as a Technique for Solvency Analysis"
[9] S.Sravan Kumar, T.RangaBabu," Emotion and Gender Recognition of Speech Signals Using SVM", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 4, Issue 3, May 2015
[10] P. Bhardwaj , S. Debbarma," A Study of Methods Involved In Voice Emotion Recognition", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 2, February 2014
[11] Björn Schuller, Gerhard Rigoll, and Manfred Lang," Hidden markov model-based speech emotion recognition", IEEE International Conference on Acoustics, Speech and Signal Processing • August 2003



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY