IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Emotion Recognition Based on Speech Using Machine Learning

N.D sonawane, A.U.Gund,P.N.Gawate,A.D Salunke,S.Kakade

Abstract – In human communication, feelings play an important role. The ability to decipher human emotions by breaking down voice is appealing in a variety of contexts. Feelings recognition can be found in a variety of places, such as the connection between PCs and people and call centres. Feeling recognition had previously relied on simple classifiers on bag-of-words models. Nonetheless, the current work on voice feeling recognition was done with the use of deep learning algorithms on static voice data. The proposed technique focuses on improving the overall precision of feeling acknowledgment during calls including artificial intelligence. The overall goal is to accurately perceive the various emotions that a single utterance conveys semantically.

.

Keywords- Speech conversion, human emotion, feature extraction, machine learning

I. Introduction

Emotions are described as conscious affect attitudes that are manifested in the expression of an emotion. A vast number of studies have focused on emotion detection via opinion mining on speech in recent years. Emotion recognition on conversations is difficult because to the inherent properties of the voice produced during calls, such as loudness, voice quality, and casual expression. Previous research has primarily focused on lexical and deep learning techniques. The quality of the emotional lexicon has a significant impact on the performance of vocabulary-based methods, and the characteristics have a significant impact on the performance of deep learning methods. As a result, we choose two of the most well-known classifiers, both of which have been utilised previously by academics in computational linguistics and natural language processing (NLP). Finally, Profile of Mood States is a psychological tool that uses text to express a four-dimensional mood state. A new methodology for producing a Profile of Mood States has been developed.

A four-dimensional mood state depiction employing 65 adjectives and a mix of emotions such as anger, happiness, sadness, and normality. Previous research focused on only one type of emotion recognition. Working with numerous categories at the same time allows us to compare performance of different emotion categorizations on the same sort of data, as well as construct a single model that can predict numerous categories at the same time.

II. LITERATURE REVIEW

The primary idea of the research [1] is to use Deep Neural Network (DNN) and knearest neighbour (k-NN) in grouping feelings from a discourse perspective that is particularly worrisome. The medical services units are primarily concerned with the framework's usage area. The foundation of this investigation has its most solid applicability in palliative care. Alarm signals are sent via cloud in the case of the most precise outcome. A large amount of raw data is obtained using a variety of accentuation techniques. The conversion of voice signals to wave structure, expression level element extraction, feeling recognition, and ready sign production via cloud are the processes to be followed.

The Teager Energy Operator (TEO) and Linear Prediction Coefficient (LPC) highlights as T-LPC include extraction are used in this paper [2] to offer a Speech Emotion Recognition (SER) framework. The presented strategies were used to perceive discourse signals that were not accurately perceived in previous SER frameworks. In this study, the Gaussian Mixture Model (GMM) classifier is used to sort the emotions in the EMO-DB data base. In comparison to the present Pitch, LPC, and LPC + Pitch inclusion based acknowledgment frameworks, the Stressed Speech Emotion Recognition (SSER) developed using the T-LPC highlight extraction approach achieved improved performance. This proposed feeling recognition framework can be used to motivate understudies by determining their enthusiastic mood with greater precision than the current ones.

Another organising model (CNN-RF) is proposed in this research [3] in light of a neural convolution network linked with an arbitrary timberland. To begin, the neural convolution network is used as a highlight extractor to separate the qualities of vocal sensations from the standardised spectrogram, and the characteristics of voice sensations are organised using an arbitrary woods order computation. The results of the test reveal that the CNN-RF model outperforms the standard CNN model. Furthermore, Nao's Record Sound Order Box was improved, and the CNN-RF model was applied to the robot.

The paper [4] constructs a perform various tasks DNN for learning assignments across numerous errands, not only by utilising massive amounts of cross-task data, but also by taking advantage of a regularisation effect that stimulates more broad depictions to assist projects in new regions. A performs a variety of tasks for representational learning that require deep brain organisation, with a focus on semantic characterization (inquiry grouping) and semantic data recovery (positioning for web search) assignments. Inquiry order and web search results should be solid. The following are focal points: Across all web search and question characterization missions, the MT-DNN consistently outperforms using solid baselines. Multiple tasks must be completed The DNN model successfully connects assignments that are as distinct as grouping and positioning. Inconveniences include the following: The investigation plan was consolidated either as a grouping or positioning project, rather than a comprehensive examination.

Show that emotion word hashtags are valid manual names of feelings in tweets in article [5]. From this feeling-marked tweet corpus, proposes an approach for generating a large lexicon of word-feeling relationships. This is the only dictionary that has true word feeling connection scores. The following are some points of interest: Using hashtagged tweets, you may get a lot of named information for any sensation that is used as a hashtag by tweeters. The hashtag emotion vocabulary performs fundamentally better than those who used the WordNet impact dictionary that was physically created. As a result, it detects characters in text. The following are some disadvantages: This paper only works with the content provided, not with equivalent words from the text.

The focus of the paper [6] is on two central NLP tasks: discourse parsing and sentiment analysis. The enhancement of three free recursive neural nets: the 1/3 web for presumption expectation and the main sub-commitments of discourse parsing, expressly structure forecast, and connection expectation. The following are some points of interest: The highlights from idle Discourse can be used to complement the presentation of a neural estimating analyzer. The individual and pre-planning approaches are significantly faster than the Multi-entrusting model. Obstacles include: Expectations for multi-sentential content are difficult to meet.

In this paper [7] they we present our discoveries on how the learning of portrayal in a huge plain vocal corpus can be utilized in a helpful manner for the acknowledgment of the feelings of language (BE). We show that the incorporation of the took in portrayals from an unattended programmed encoder into a CNN-based feeling classifier improves acknowledgment exactness.

In this paper [8], He proposed another model for the nonstop acknowledgment of feelings from language. This model, which has been prepared from one finish to the next, is made out of a convolutional neural organization (CNN), which extricates the attributes of the natural sign and stacks a momentary long haul 2-layer memory (LSTM), for Consider the logical data in the information.

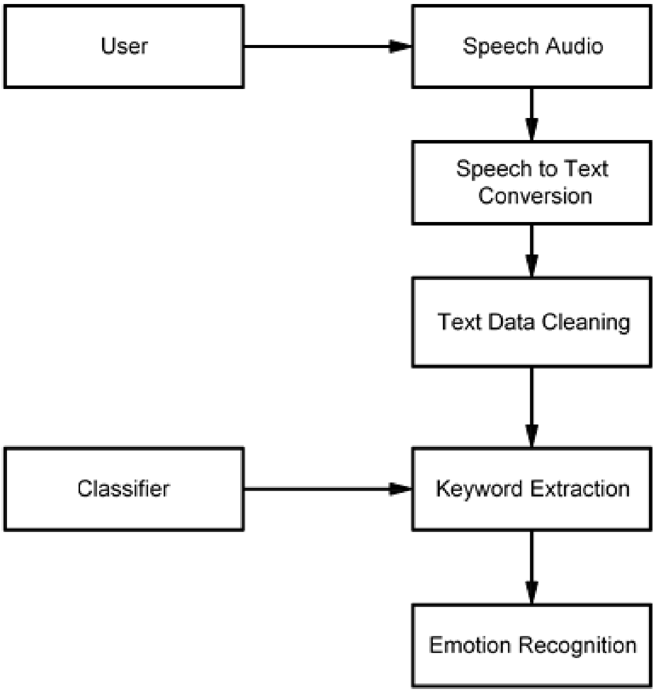
This paper [9] propose to incorporate the consideration system into profound intermittent neural organization models for discourse feeling acknowledgment. This depends on the instinct that it is useful to underscore the expressive piece of the discourse signal for feeling acknowledgment. By presenting consideration instrument, we cause the framework to figure out how to zero in on the more hearty or useful portions in the info signal. The proposed incorporation of consideration instrument on top of the gauge profound RNN model accomplishes 46.3% UA review rate.

In this paper [10] they utilize 13 MFCC (Mel Frequency Cepstral Coefficient) with 13 speed and 13 increasing speed segment as highlights and a CNN (Convolution Neural Network) and LSTM (Long Short Term Memory) based methodology for arrangement. We picked Berlin Emotional Speech dataset (EmoDB) for grouping reason. We have roughly 80% of precision on test information.

III. Proposed Approach

Component extraction algorithms extract text from human discourse using a discourse transformation library. Human Mood States is a psychological tool for assessing a person's temperament. It refers to a set of 65 descriptive words that the subject rates on a five-point scale. Every descriptive word contributes to one of the four categories. With the exception of loose and effective, whose commitments to their particular classes are negative, the higher the modifier's score, the more it contributes to the overall score for its classification. Disposition states combines these assessments into a four-dimensional temperament state depiction that includes four categories: indignation, joy, tragedy, and typical. In contrast to the original structure, we eliminated the modifier blue because it is rarely associated with a sentiment.

A. System Architecture



B. Algorithm:

1. Hidden Markov Model (HMM) algorithm for speech recognition:

A HMM is characterized by 3 matrices viz., A, B and PI.

A - Transition Probability matrix $(N \times N)$

B - Observation symbol Probability Distribution matrix $(N \times M)$

PI - Initial State Distribution matrix $(N \times 1)$

Where, N =Number of states in the HMM

M = Number of Observation symbols

After can apply HMM for speech recognition by using following steps:

1. Recursive procedures like forward and Backward Procedures exist which can compute P (O|L), probability of observation sequence.

Forward Procedure:

Initialization:

$\alpha_1(i) = \pi_i b_i o_1, \ 1 \leq i \leq N$

Induction

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(o_{t+1}),$$

$$1 \leq t \leq T-1, 1 \leq j \leq N$$

Termination

$$P(O|\lambda)\sum_{i=1}^{N}\alpha_T(i)$$

Backward Procedure:
Initialization:
$\beta_T(i) = 1,\ 1 \le i \le N$
Induction

$$\beta_T(i) = \sum_{j=1}^{N} a_{ij}\, b_j(o_{t+1})\beta_{t+1}(j),$$

$$T-1 \le t \le 1, 1 \le i \le N$$

Termination

$$P(O|\lambda)\sum_{i=1}^{N}\alpha_T(i)$$

2. The state occupation probability $t(sj)$ is the probability of occupying state sj at time t given the sequence of observations

$$O_1, O_2, \ldots, O_N.$$

3. Baum-welch algorithm for parameter re-estimation.

2. MMLDA Algorithm for summarization

Steps:

1. For the topic T, draw $\varphi^{TG} \sim Dir(\lambda^{TG})$ and $\varphi^{VG} \sim Dir(\lambda^{VG})$ denote the general textual distribution and visual distribution, respectively. $Dir(\cdot)$ is the Dirichlet distribution. Then draw $\phi^Z \sim Dir(\beta^Z)$, which indicates the distribution of subtopics over the microblog collection corresponding to T.
2. For each subtopic, draw $\varphi_K^{TS} \sim Dir(\lambda^{TS})$ and $\varphi_K^{VS} \sim Dir(\lambda^{VS}),\ k \in \{1, 2, \ldots, K\}$, correspond to the specific textual distribution and visual distribution.
3. For each microblog M_i, draw $Z_i \sim Multi(\phi^Z)$, corresponds to the subtopic assignment for Mi. Multi(·) denotes the Multinomial distribution. Then draw $\phi_i^R \sim Dir(\beta^R)$ indicates the general-specific textual word distribution of M_i. Similarly, draw $\phi_i^Q \sim Dir(\beta^Q)$ indicates that for visual words.
4. For each textual word position of M_i, draw a variable $R_{ij} \sim Multi(\phi_i^R)$:
    - If R_{ij} indicates General, then draw a word $W_{ij} \sim Multi(\varphi^{TG})$.
    - If R_{ij} indicates Specific, draw a word W_{ij} from the Z_i-th specific distribution $W_{ij} \sim Multi(\varphi_{Z_i}^{TS})$
5. The generation of visual words is similarly done as in step 4.

IV. Results and Discussion

The experimental result evaluation, we have notation as follows:

TP: True positive (correctly predicted number of instance)

FP: False positive (incorrectly predicted number of instance),

TN: True negative (correctly predicted the number of instances as not required)

FN false negative (incorrectly predicted the number of instances as not required),

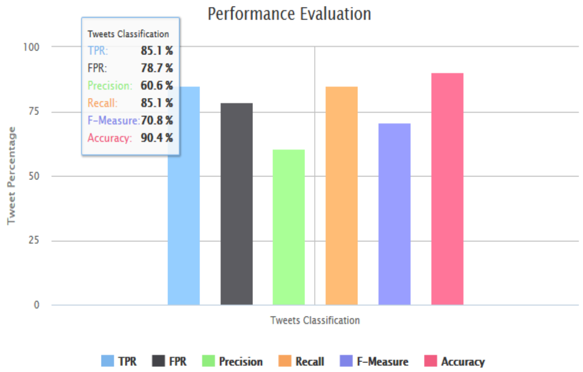
On the basis of this parameter, we can calculate four measurements

Accuracy = TP+TN÷TP+FP+TN+FN

Precision = TP ÷TP+FP

Recall= TP÷TP+FN

F1-Measure = 2×Precision×Recall ÷Precision+ Recall.



Parameters	Percentage
TPR	85.1
FPR	78.7
Precision	60.6
Recall	85.1
F-Measure	78.8
Accuracy	94.4

Conclusion

Understanding a person's temperament or state through speech is a new notion whose use is undeniable and which will find applications in a range of sectors, from therapeutic to data enhancement. This study uses a novel methodology called Profile of Mood States (POMS), which employs multinomial innocent Bayes to depict four-dimensional temperament states utilising 65 modifiers and a mix of emotional classifications including cheery, unhappy, outrage, and normal.

References

[1] K.Tarunika , R.B Pradeeba , P.Aruna" Applying Machine Learning Techniques for Speech Emotion Recognition" ICCCNT 2018.

[2] Surekha Reddy B, T. Kishore Kumar" Emotion Recognition of Stressed Speech using Teager Energy and Linear Prediction Features" 2018 IEEE 18th International Conference on Advanced Learning Technologies

[3] Li Zheng, Qiao Li2, Hua Ban , Shuhua Liu1" Speech Emotion Recognition Based on Convolution Neural Network combined with Random Forest"IEEE 2018

[4] O. Irsoy and C. Cardie, "Opinion Mining with Deep Recurrent Neural Networks," in Proc. of the Conf. on Empirical Methods in Natural Language Processing. ACL, 2014, pp. 720–728.

[5] S. M. Mohammad and S. Kiritchenko, "Using Hashtags to Capture Fine Emotion Categories from Tweets," Computational Intelligence, vol. 31, no. 2, pp. 301–326, 2015.

[6] B. Nejat, G. Carenini, and R. Ng, "Exploring Joint Neural Model for Sentence Level Discourse Parsing and Sentiment Analysis," Proc. of the SIGDIAL 2017 Conf., no. August, pp. 289–298, 2017.

[7] Michael Neumann, Ngoc Thang Vu," IMPROVING SPEECH EMOTION RECOGNITION WITH UNSUPERV"IEEE 2019.

[8] Panagiotis Tzirakis, Jiehao Zhang, Bjorn W. Schuller "END-TO-END SPEECH EMOTION RECOGNITION USING DEEP NEURAL NETWORKS"IEEE 2018

[9] Po-Wei Hsiao and Chia-Ping Chen" EFFECTIVE ATTENTION MECHANISM IN DYNAMIC MODELS FOR SPEECH EMOTION RECOGNITION"IEEE 2018

[10] Saikat Basu, Jaybrata Chakraborty, Md. Aftabuddin" Emotion Recognition from Speech using Convolutional Neural Network with Recurrent Neural Network Architecture" International Conference on Communication and Electronics Systems (ICCES 2017)

1)Nayana Sonawane
Email address-2nayna@gmail.com
2)Ashwini.U.Gund

Email address- ashwinigund634@gmail.com

3)Puja Gawate

Email address-Pvgaikwad0304@gmail.com

4) Akshata Salunke

Email address -akshatasalunke2000@gmail.com

5) Sumeet Kakade

Email address-kakadesumeet512@gmail.com