



EMOTION DETECTION THROUGH AUDIO USING MACHINE LEARNING

Computer Science

V. Kranthi Sai
Reddy

Student, Sreenidhi Institute of Science and Technology, Hyderabad 500039

ABSTRACT

Emotion recognition through Audio signal is a recent research topic in the Human Computer Interaction. The demand was risen for increasing communication interface between the humans and digital media. Many researchers are currently working in order to improve their accuracy. But still there is a lack of complete system which can recognize emotions from speech. In order to make the human and the digital machine interaction more natural, the computer should be able to recognize emotional states in the same way as human. The voice of a person plays an important role in analyzing the people. From voice of the person we are detecting the emotion of the person. Emotion detection has its importance in forensics, games, in security purposes and of course in our day to day life. To detect the emotion pitch, speaking rate and energy are taken as features and AdaBoost with C4.5 ML algorithm is taken for classification and achieved 93.12% accuracy.

KEYWORDS

Emotion Detection, AdaBoost, C4.5, classification

INTRODUCTION

Emotion detection from speech or audio is a challenging problem in audio signal processing. Lot of information like: age, gender, emotion, person and action can be estimated from a audio signal, emotion detection is one of them. It is completely known fact that the human's speech encloses linguistic content, identity as well as the emotion of speaker. Emotion plays as one of the significant role in daily interpersonal human interactions. It is essential to our rational as well as intelligent decisions. It will help us to match and understand the feelings of others by conveying our feelings and giving feedback to others and speech is relevant communicational channel enriched with emotions.

Human-Technology interface is significant in both the quantitative and qualitative terms. In order to improve the experience and to enhance the features of several hi-performance and hi-tech applications and also features like emotion detection and emotional-based feedback of the machines are challenging directions to the research grouped under the umbrella-term 'affective computing'. These systems can be helpful in detecting the customers' emotion, medical entertainment, and crime detection, robotics voice and may other cases. Speech communication contains paralinguistic information of the speaker. Although an enormous effort is invested in recognizing the emotions of a person from speech but still much research is needed.

There are already several significant results and applications of affective computing that are available for robots, computers and mobile devices. This means that machines already recognize some fundamental human emotions but also there is still a long way to go. These systems can be useful in detecting the customers' emotion, medical entertainment, and crime detection, robotics voice and may other cases. Speech communication contains a paralinguistic information of the speaker. Although enormous efforts are invested in recognizing the emotions from speech or audio but still much research is needed.

Emotions are universal but their understanding, interpretation and reflections are particular and partly culturally specific. Based on the art survey of results in emotion detection, we decided to implement the emotion detection from voice, as most appropriate in the context of application intended. There are many applications to detect the emotion of the persons like in audio surveillance, web-based E-learning, and commercial applications, clinical studies, and entertainment etc. Emotion identification can be used as a voice tag in different database access systems. This voice tag is used in telephony shopping, and ATM machine as a password for accessing that particular account. We aim to analyze the role of fundamental frequency and other features such as pitch, energy and speaking rate are extracted for detecting the emotion of the person.

METHODOLOGY

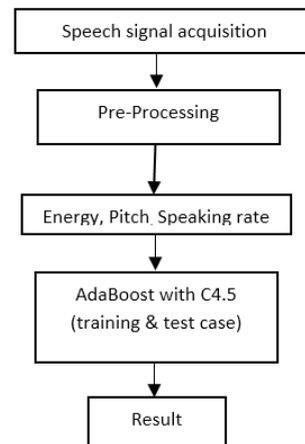


Fig 1: Proposed system for Emotion Detection

WORK DONE

- Speech signal acquisition: The speech signals were captured from 300 people including both male and female in the age group of 20-30 years with the help of microphones. All the voice samples for emotion detection were recorded under circumstances that means samples may contains the noise like of fan or any other common noise. For emotion detection people were told to speak in four different emotions i.e., 'happy', 'normal', 'sad' and 'angry'. The sampling has been done while recording the voice sample as well. Sampling is described as one of the formations of discrete signal from the continuous signal. The speech signals captured were sampled at 44 kHz. Praat is used for recording for audio signals.

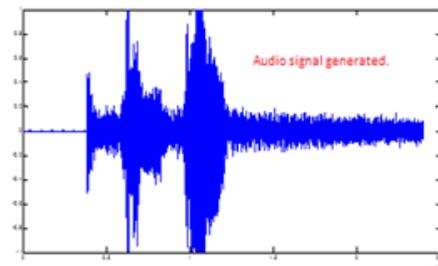


Fig 2: Recording the Audio Signals

- Pre-Processing: Pre-Processing is done to remove the noise present in the collected samples. This includes pre-emphasis and windowing. Pre-emphasis process is used to remove the noise

from the captured signal and also spectral subtraction method has been used to remove the noise and further hamming window is used for hamming window is used for windowing.

- c. Feature extraction: Speaking Rate, pitch and energy are extracted for emotion detection. Features were extracted from the voice signals using praat tool. Pitch is one of the most essential components of emotion recognition through audio signal. It defines the rate of vibration of speaker's vocal cord. Although different sub features like fundamental frequency, pitch, energy and speaking rate are used.
- d. Classification: The feature vector is constructed using the above-mentioned features which were fed to ML algorithms for recognition. AdaBoost with C4.5 is used for emotion detection. Classification is done by using the weka which is developed at New Zealand based University of Waikato. Weka is an open simulator and has a java-based implementation of the ML algorithms and researchers use it extensively.

C4.5 & ADABOOST

C4.5 is a machine learning algorithm used to generate a decision tree and developed by Ross Quinlan. C4.5 is an extension of the Quinlan's earlier ID3 algorithm. The decision trees which are generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

AdaBoost stands for Adaptive Boosting, it is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire. It can be used as conjunction with many other types of learning algorithms to improve performance. AdaBoost is mainly used to boost the performance of the decision trees on binary classification problems. AdaBoost because it is most used for classification rather than regression. AdaBoost is used to boost the performance of any machine learning algorithm. These are the models that achieve accuracy just above random chance on a classification problem. In our project AdaBoost is used to boost the performance of C4.5

EXPERIMENTAL RESULTS

The project was carried out by taking voice samples from 300 male and female speakers of age group 20-30 years. Voice samples were taken under normal circumstances for emotion detection. Speakers were told to speak in four different emotions 'happy', 'normal', 'sad' and 'angry' mood for emotion detection. Praat tool was used to collect the voice samples. From voice samples pitch, energy, and speaking rate were extracted for emotion detection. For emotion recognition machine learning algorithm named as MLP is used for classification and AdaBoost with C4.5 was used.

For emotion detection MLP did not perform well as its recognition rate was low i.e., 51.2% as MLP performs well if the number of input units will be less but in emotion detection, we have more number of input units. Then we applied the C4.5 algorithm as it is fast in performing the classification task and got 76.25% recognition rate. To enhance the performance of C4.5 algorithm we used the AdaBoost algorithm as it is a boosting algorithm that can be used to improve the performance and we got very good results with average accuracy of 93.12%. Comparison is shown in fig 3.

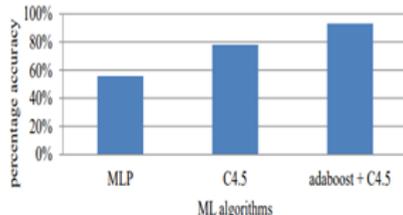


Fig 3: Comparison of the ML algorithms for emotion detection

The percentage of the emotional accuracy for both male and female shown in the below figure

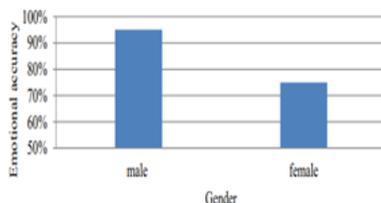


Fig 4: Emotional Accuracy for both male and female

As we analyzed emotional accuracy varies with the gender so this bar chat shows that the accuracy in case of males is higher than the females. It is difficult to detect the emotion of the females are compared to males as sometimes females express in two different emotions in same manner, they react exactly at the two different moments like they speak loudly in happy mood but also in sad mood.

Moreover, sometimes they will speak slowly in angry mood too and in sad mood. While in case of males they mostly react differently or we can also say they express different emotions in different manners. That's why it becomes easier to detect the emotion of the males.

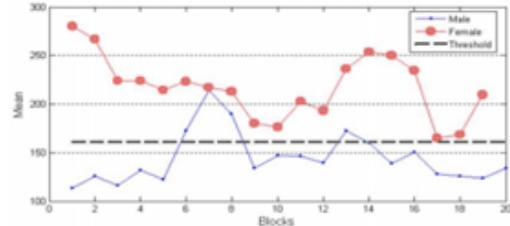


Fig 5: Represents average pitch frequencies of male and female

CONCLUSION

Emotion recognition from a speech signal is useful for various applications like in the field of human machine interaction. In order to develop this type of system emotion recognition is gaining good demand. The work presented in paper has proposed a feature extraction and recognition system to recognize emotion of the persons. The features that are taken in our proposed work comprise pitch, energy and speaking rate for emotion detection. For classification MLP is used and AdaBoost with C4.5 is used for emotion detection. Algorithms were used and got 93.12% for emotion detection. Moreover, it is also concluded that the emotion varies with gender, from experimentation, it is difficult to detect emotion of females as compared to males.

FUTURE WORK

This project can be expended further in future by increasing the dataset as we have taken the voice samples of 300 people, one can take the greater number of voice samples. Further samples can be taken in the vacuum, as the noise affects the accuracy of the emotion detection, it may give the better results. Moreover, different feature set can be included as we have taken pitch, energy and speaking rate for emotion detection, in future more features can be added like MFCC coefficients and LPC coefficient etc. This work can be further expended by using the other emotions like boredom etc. as we have worked only on four emotions happy, normal, sad and angry.

REFERENCES:

1. P. Ekman, An argument for basic emotions, *Cognition and Emotion* 6 (1992) 169-200.
2. Shashidhar G. Koolagudi, Ramu reddy, Jainath Yadav, K.Sreenivasa Rao. "IITKGP-SEHSC:Hindi speech corpus for emotion analysis." *IEEE* (2011).
3. Agrawal, S. S., Prakash, N., & Jain, A. (2010). Transformation of emotion based on acoustic features of intonation patterns for Hindi speech. *IJCSNS International Journal of Computer Science and Network Security*, 10(9), 198-205.
4. Albornoz, Enrique M., Diego H. Milone, and Hugo L. Rufiner. "Feature extraction based on bio-inspired model for robust emotion recognition." *Soft Computing* (2016): 1-14.
5. Rajiv Kumar and Kiran Kumar Ravulakollu, "Handwritten Devnagari Digit Recognition: Benchmarking on new dataset", *Journal of Theoretical and applied Information Technology* (E-ISSN 1817-3195 / ISSN 1992-8645), 2014, vol. 60, no. 3, 2014, pp. 543-555.
6. X. rathina, K.M. mehata, M. Ponnavaikko," basic analysis on prosodic features in emotional speech", *international journal of computer science, engineering and applications* vol.2, No. 4, august 2012, PP99-107,2012.
7. Bahuguna, Sushma, and Y. P. Raiwani. "Study of Speaker's Emotion Identification for Hindi speech." *International Journal on Computer Science and Engineering* 5, no. 7 (2013): 629-634.
8. WEKA 3: Data Mining With Open Source Machine Learning Software in JAVA.
9. Dipti D. Joshi1, Prof. M. B. Zalte," Speech Emotion Recognition: A Review", *IOSR Journal of Electronics and Communication Engineering*, Volume 4, Issue 4 (Jan. - Feb. 2013), PP34-37,2012.
10. Mehmet Cenk Sezgin, Bilge Günsel* and Gunes Karabulut Kurt, "Perceptual audio features for emotion detection", *EURASIP Journal on Audio, Speech, and Music Processing*, 2012.