Gender with Emotion Recognition Using Machine Learning

K. Keerthika[1], K. Monika[2], M. Monisha[3], H. Nandhini[4]

[1, 2, 3, 4]Student, Computer Science and Engineering, Kingston Engineering College, Vellore, TamilNadu, India

Abstract: ***In this paper, we study how speech features' numbers and statistical values impact recognition accuracy of emotions present in speech. With Gaussian Mixture Model (GMM), we identify two effective features, namely Mel Frequency Cepstrum Coefficients (MFCCs) extracted directly from speech signal. Using GMM supervector formed by values of MFCCs, delta MFCCs and ACFC, we conduct experiments with Berlin emotional database considering six previously proposed emotions: anger, disgust, fear, happy, neutral and sad. Our method achieve emotion recognition rate of 74.45%, significantly better than 59.00% achieved previously. To prove the broad applicability of our method, we also conduct experiments considering a gender and different set of emotions: anger, boredom, fear, happy, neutral and sad.Our emotion recognition rate of 75.00% is again better than71.00% of the method of hidden Markov model with MFCC, delta MFCC, cepstral coefficient and speech energy.***

Keywords: Accuracy, Gender recognition, Emotion detection

1. Introduction

Research in intelligent healthcare has significantly expanded during the past few years. When caring for people who have lost the ability to convey intentions directly, corporeal machine interfaces are one way of restoring communication. Despite remarkable achievements in corporeal machine interfaces, the recognition of patients' emotions remains a critical challenge. Emotion recognition is a crucial component of emotion interaction and the first step towards understand the feelings of patients. Brain imaging technology has allowed researchers to non-invasively detect brain activity during emotional processing like facial expression, voice, body posture, etc. emotion-related neural activity is difficult to hide and is thus theoretically more reliable for emotion recognition. Further, compared to physiological-signal-based emotion recognition (cutaneogalvanic, electromyograph, electrocardiograph, etc.), neural activity provides better information for emotion recognition due to its higher specificity for different types of emotions. Electroencephalograph (EEG) and functional magnetic resonance imaging (fMRI) have been used to record emotion-related neural activity. These studies have shown the great potential of neural-signal-based emotion recognition.

1) We quantify the accuracy instability of several fNIRS classification schemes for emotion recognition when using training data collected 3 weeks earlier.
2) We analyze several possible contributing factors to data instability and accuracy instability.
3) We propose and evaluate a novel feature selection method for fNIRS emotion recognition to mitigate inter-session accuracy instability.

Artificial Intelligence

Artificial intelligence is a branch of computer science that aims to create intelligent machines. It has become an essential part of the technology industry.

Research associated with artificial intelligence is highly technical and specialized. The core problems of artificial intelligence include programming computers for certain traits such as:

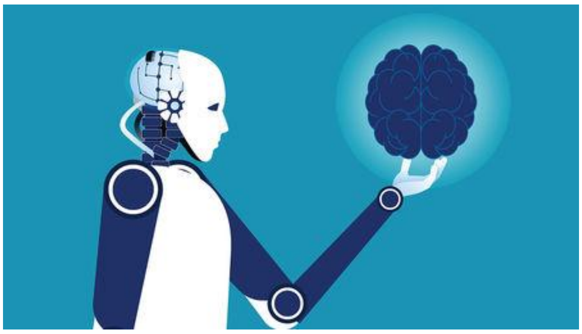
- Knowledge
- Reasoning
- Problem solving
- Perception
- Learning
- Planning
- Ability to manipulate and move objects

Knowledge engineering is a core part of AI research. Machines can often act and react like humans only if they have abundant information relating to the world. Artificial intelligence must have access to objects, categories, properties and relations between all of them to implement knowledge engineering. Initiating common sense, reasoning and problem-solving power in machines is a difficult and tedious task.

Machine learning is also a core part of AI. Learning without any kind of supervision requires an ability to identify patterns in streams of inputs, whereas learning with adequate supervision involves classification and numerical regressions. Classification determines the category an object belongs to and regression deals with obtaining a set of numerical input or output examples, thereby discovering functions enabling the generation of suitable outputs from respective inputs. Mathematical analysis of machine learning algorithms and their performance is a well-defined branch of theoretical computer science often referred to as computational learning theory.

Machine perception deals with the capability to use sensory inputs to deduce the different aspects of the world, while computer vision is the power to analyze visual inputs with a few sub-problems such as facial, object and gesture recognition.

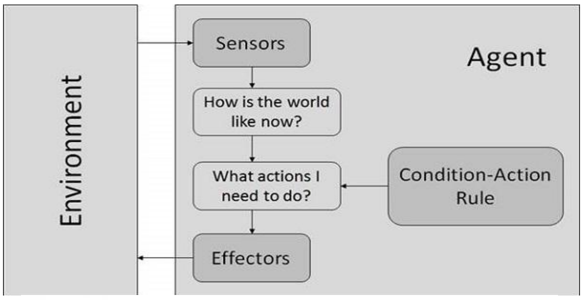
Robotics is also a major field related to AI. Robots require intelligence to handle tasks such as object manipulation and navigation, along with sub-problems of localization, motion planning and mapping.

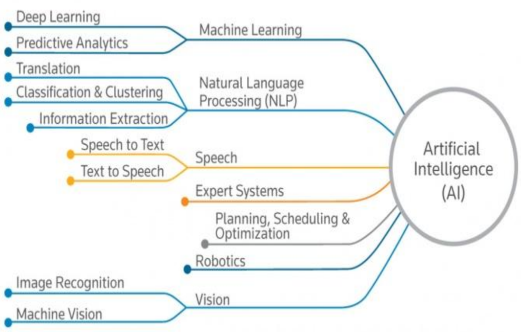


1.1 Objective

1) To classify the gender based on speech recognition.
2) To identity the emotion of the speaker.
3) And to show the accuracy between the existing system and proposed system.

1.4 Structure of AI





1.5 Feasibility Study

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- Economical Feasibility
- Technical Feasibility
- Social Feasibility

Economical Feasibility
This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical Feasibility
This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social Feasibility
The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

2. Problem Statement

The problem that we want to solve is to reduce the noise form the give human voice and also recognize the emotion of the speaker

3. System Design

3.1 System Architecture

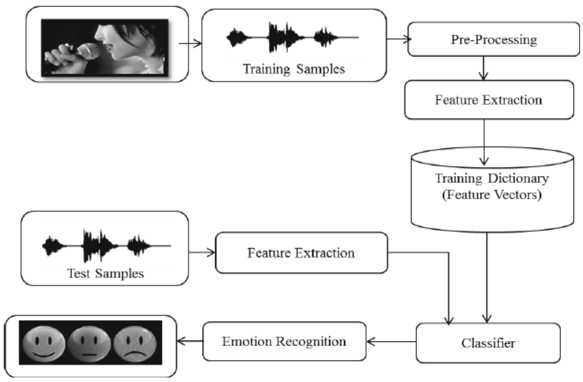


Figure 3.1: System Architecture

- Mel-frequency cepstral coefficients (MFCCs)
- Hidden Markov Model (HMM)
- Multi Support vector machine (SVM)

3.2 MEL-Frequency Cepstral Coefficients (MFCCS)

Mel-frequency cepstral coefficients (**MFCCs**) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

- MFCCs are commonly derived as follows:
- Take the Fourier transform of (a windowed excerpt of) a signal.
- Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
- Take the logs of the powers at each of the mel frequencies.
- Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
- The MFCCs are the amplitudes of the resulting spectrum.

3.3 Hidden Markov Model (HMM)

- Hidden Markov models are especially known for their application in reinforcement learning and temporal pattern recognition as speech, handwriting, gesture recognition ,part-of-speech tagging,musical score following partial discharges and bio-informatics.
- A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other.
- Recently, hidden Markov models have been generalized to pairwise Markov models and triplet Markov models which allow consideration of more complex data structures and the modeling of nonstationary data.

3.4 Multi Support Vector Machine (SVM):

- In machine learning, **support vector machines** (**SVMs**, also **support vector networks**[1]) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.
- Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier(although methods such as Platt scaling exist to use SVM in a probabilistic classification setting).
- An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.
- New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

4. Execution Setup

When the user submits the voice note the system will remove the unwanted noise from the voice note . once the unwanted noises are removed .the voice is changed in to the frequency .the system will recognize whether the voice belong to male are female in addition it also recognize the emotion of the speaker eg: Angry ,Sad, Happy etc.

5. Conclusion

Speech emotion recognition systems based on the several classifiers is illustrated. The important issues in speech emotion recognition system are the signal processing unit in which appropriate features are extracted from available speech signal and another is a classifier which recognizes emotions from the speech signal. The average accuracy of the most of the classifiers for speaker independent system is less than that for the speaker dependent.

Automatic emotion recognitions from the human speech are increasing now a day because it results in the better interactions between human and machine. To improve the emotion recognition process, combinations of the given methods can be derived. Also by extracting more effective features of speech, accuracy of the speech emotion recognition system can be enhanced.

References

[1] D. Reynolds, "Gaussian Mixture Models", Encyclopedia of Biometric Recognition, Springer, Feb. 2008.
[2] D. A.Reynolds and R. C. Rose,"Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", IEEE transactions on speech and audio processing, vol. 3, pp. 72-83, January 1995.
[3] Jyh-Shing Roger Jang, "Data Clustering and Pattern Recognition", available at the links for on-line courses at the author's homepage at http://mirlab.org/jang.
[4] Moataz El Ayadi, Mohamed S.Kamel, Fakhri Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases", Pattern Recognition, Volume 44, Issue 3, pp. 572-587, March 2011.
[5] S. Bou-Ghazale, J. Hansen, "A comparative study of traditional and newly proposed features for recognition of speech under stress", IEEE Trans. Speech Audio Process, 8 (4) (2000) 429-442
[6] M. E. Ayadi, M. S. Kamel, F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases", Pattern Recognition 44, PP.572-587, 2011.
[7] I. Chiriacescu, "Automatic Emotion Analysis Based On Speech", M.Sc. THESIS Delft University of Technology, 2009.
[8] T. Vogt, E. Andre and J. Wagner, "Automatic Recognition of Emotions from Speech: A review of the literature and recommendations for practical realization", LNCS 4868, PP.75-91, 2008.
[9] S. Emerich, E. Lupu, A. Apatean, "Emotions Recognitions by Speech and Facial Expressions Analysis", 17th European Signal Processing Conference, 2009.
[10] A. Nogueiras, A. Moreno, A. Bonafonte, Jose B. Marino, "Speech Emotion Recognition Using Hidden Markov Model", Eurospeech, 2001.
[11] P.Shen, Z. Changjun, X. Chen, "Automatic Speech Emotion Recognition Using Support Vector Machine", International Conference On Electronic And

Mechanical Engineering And Information Technology, 2011.

[12] D. Ververidis and C. Kotropoulos, "Emotional Speech Recognition: Resources, Features and Methods", Elsevier Speech communication, vol. 48, no. 9, pp. 1162-1181, September, 2006.

[13] Z. Ciota, "Feature Extraction of Spoken Dialogs for Emotion Detection", ICSP, 2006.

[14] E. Bozkurt, E, Erzin, C. E. Erdem, A. Tanju Erdem, "Formant Position Based Weighted Spectral Features for Emotion Recognition", Science Direct Speech Communication, 2011.

[15] C. M. Lee, S. S. Narayanan, "Towards detecting emotions in spoken dialogs", IEEE transactions on speech and audio processing, Vol. 13, No. 2, March 2005