Discovery of Human Emotion using Fuzzy based Cognitive Model

Dr. A. Akila
Associate Professor
&
Dr. R. Parameswari
Associate Professor
Department of Computer Science
School of Computing Sciences
Vels Institute of Science,Technology & Advanced Studies (VISTAS)
(Deemed University)
Chennai, Tamil Nadu, India

Abstract

The emotion of a human could be identified using Speech, Image and Question and answer session. Also, the emotion in speech is identified using the pitch and intensity. The emotion identification with image is done using Support Vector Machine. The present chapter envisages into an intelligent system, which is designed to understand human emotions more precisely speech emotion identification and intends to generate actions via cognitive system. It has mainly focused on developing an online incremental learning system of human emotions using Takagi-Sugeno (TS) Fuzzy model. The main objective of this system is to detect whether the observed emotion needs a new corresponding multi-model action to be generated or it can be attributed to one of the existing actions in memory. The multi-model consists of voice input, facial expression. The combined results have been classified using TS Fuzzy Model.

Keywords: Mel Frequency Cepstral Coefficient, Speech, Emotion, Identification, TS Fuzzy Model, Pitch, Intensity, Cognitive, Image.

Introduction

Emotion plays a significant role in daily interpersonal human interactions. This is essential to our rationale to make intelligent decisions. Also, it helps us to match and understand the feelings of others by conveying our feelings and giving feedback to others. The research studies have revealed the powerful role that emotion play in shaping human social interaction. It considers about the thinking of human brain and how the nervous function works. The main domain is Cognitive Science, which means study of thought and mind. Moreover, there two modules used namely,

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

speech recognition and facial expression recognition. In speech recognition, we use a MFCC [1] [2] (Mel Frequency Cepstral Coefficient). It is a technique, which takes voice sample as input and one can use audacity software for voice record in .WAV format. Furthermore, if one wants to convert an .MP3 file to .WAV format, the audacity could be used. This in turn helps to change the voice file format from MP3 and WAV to WAV and MP3 respectively. In facial recognition, one can include many kinds of emotions i.e., anger, sad, happy, surprised, neutral and disgust. For instance, if one gives an image into a process, it will check and find out the kind of emotions that are hiding in the face and showing it. Keeping these aforementioned aspects, this research chapter attempts to find out the emotions through speech and facial expressions.

Literature Review

Speech Based Emotion Recognition

The speech recognition system is based on Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) from which human speeches have been decoded into signals for digital processing. The endpoint detection was applied to remove unvoiced area between segment words, and then the features were extracted by Linear Predictive Coding (LPC), MFCC and Gamma Tone Cepstral Coefficient (GTCC). The phoneme model has been built from speech signals in the training database. HMM evaluation was performed to get the recognized word. Then words were composed to get the sentence in text. Dynamic Time Warping (DTW) was used for clustering the feature vectors extracted from Linear Predictive Coding (LPC), MFCC and Gamma Tone Cepstral Coefficient (GTCC). In furtherance, the spoken sentence was represented as sequence of independent acoustic phonetic units. Thus, Hidden Markov Model (HMM) was used to encode the temporal evolution of the extracted features. Gaussian distributions were used to measure variations in speaker, accent and pronunciations.

The recognition of hearing-impaired speech has been carried out by the use of Hidden Markov Model with LPC, MFCC and Perceptual Linear Predictive (PLP) features. The mixture values in HMM was selected randomly according to the number of speakers. The isolated digits in Tamil were taken as input and the performance was evaluated using accuracy and speed. The effect of fixing the appropriate number of states and the number of mixtures in the accuracy of the connected word HMM system is a major constraint. The data considered was the connected digits taken from deaf and hard hearing speech. A phoneme model was used in this work. The different features like LPC, PLP and MFCC were considered and the accuracy of the system was analyzed with 3 to 10 number of mixtures and 3 to 6 number of states. In human life emotions play important but it is very difficult to predict. Different methods are discovered for emotions recognition like SER, HMI AND MFCC.SER system identifies emotions on paralinguistic basis. Human

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

Machine Interaction derives major motivation for the work [3] [4]. MFCC coefficients in the feature vector for identifying the paralinguistic content. The core aim of emotion recognition system is to enable Human-Computer Interaction (HCI) and MFCC is one of the spectral features, wherein SVM is used for classification. Hence, to solve multi-class problems numerous single stage SCM's are used. In the last two decades, speech signals, as become one of the most natural media of human communication. Furthermore, it has been developed for automatically identifying human emotions from speech signals, which is called speech emotion recognition. The two typical deep learning methods are Deep Neural Networks (DNN), Deep Convolutional Neural Networks (DCNN).

Major Obstacles of Emotion Recognition

- Emotions are subjective, people would interpret it differently. It is hard to define the notion of emotions.
- Annotating an audio recording is challenging. Should we label a single word, sentence or a whole conversation? How many emotions should we define to recognize?
- Collecting data is complex. There are lots of audio data can be achieved from films or news. However, both of them are biased since news reporting has to be neutral and actors' emotions are imitated. It is hard to look for neutral audio recording without any bias.
- Labeling data require high human and time cost. Unlike drawing a bounding box on an image, it requires trained personnel to listen to the whole audio recording, analysis it and give an annotation. The annotation result has to be evaluated by multiple individuals due to its subjectivity.

Human Emotion Recognition Using Fuzzy Based Cognitive Model

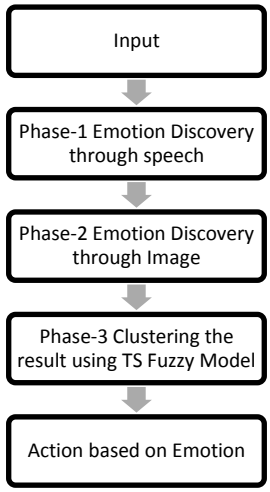

Fig.1 Human Emotion Using Fuzzy Based Cognitive Model

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

Phase 1 - Emotion Discovery Through Speech

In this section we consider about the speech techniques. In Speech signal, the voices are different from one person to another person not all the person voice are same at all time.Here we seen about the what kind of emotions are have humans vocal system [5].

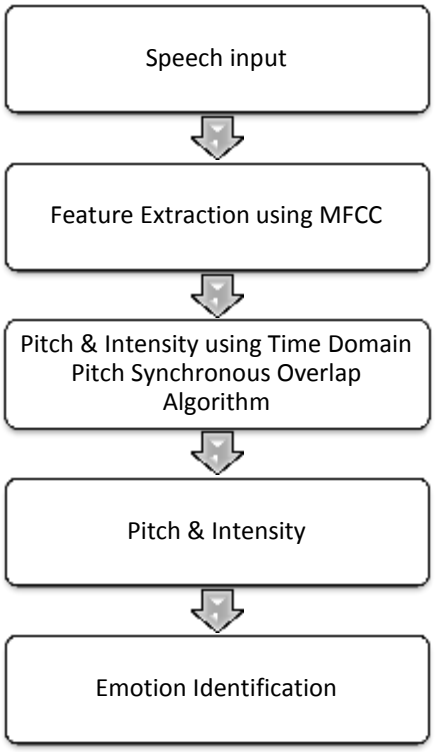

Fig.2 Speech Based Emotion Identification

The Emotions are as follows:

- Sad
- Happy
- Surprised
- Neutral
- Angry

So, herewe seen about this kinds of emotions.To find out these emotions using mffcc algorithm. MFCC algorithm means Mel Frequency Cepstral Coeffcient algorithm [6][7][8] are based on the known variations of the human ear's critical bandwidths with frequencies which are below a 1000Hz [9][10][11]. MFCC is a speech signal algorithm [12][13].

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

The energy is the first coefficient in the MFCC and the energy for each frame is manipulated[14]. The pitch is calculated and pitch range is used to identify the emotion.

$$Pitch = \frac{Energy}{Time} \quad (1)$$

$$Time = \frac{1}{frequency} \quad (2)$$

Pitch & Intensity

Pitch

Pitch depends on the frequency of a sound wave. Frequency is the number of wavelengths that fit into one unit of time. Remember that a wave length is equal to one compression and one rarefaction [15].

Pitch Range

In music, the range or chromatic range, of a musical instrument is the distance from the lowest to the highest pitch it can play. For a singing voice, the equivalent is vocal range. The range of a musical part is the distance between its lowest and highest note.

The different speech signals with different emotions were recorded and the 20 speech signals were used to train the emotion recognizer and the range of pitch for each emotion has been found.

Table: 1 Pitch Range of Different Emotions

Emotions	Pitch Range (in Hertz)
Sad	200-250
Happy	370-600
Surprised	300-350
Neutral	125-200
Angry	75-125

Phase 2 - Emotions Discovery Through Image

In this phase we seen about the image reactions and then to find out what kind of emotions having there in the face [16] [17] [18].Here we use a SVM based algorithm.SVM means support vector machine, which is fundamentally a binary classification algorithm [19]. It falls under the umbrella of machine learning. Image processing on the other hand deals primarily with manipulation of images [20] [14]. For example,image filtering,where an input image is passed through a laplacian filter to be sharpened.

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

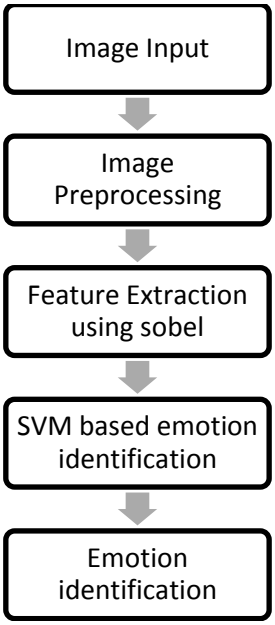

Fig.3 Image Based Emotion Identification

The image will be preprocessed to remove the noise in the image. The Feature is extracted from the image and using Support Vector Machine, the emotion is recognized. If one want to relate the two, an SVM might be used to perform image classification. For instance, given an input image, the classification task is to decide whether an image is a cat or dog. The image, before being input into the SVM might have gone through some image processing filters so that some features might be extracted such as edges, color and shape [22] [23] [24].

Feature Extraction using Sobel

The Sobel filter is used for edge detection. It works by calculating the gradient of image intensity at each pixel within the image. It finds the direction of the largest increase from light to dark and the rate of change in that direction. It works by detecting discontinuities in brightness. Edge detection is used for image segmentation and data extraction in areas such as image processing, computer vision, and machine vision. Common edge detection algorithms include Sobel, Canny, Prewitt, Roberts, and fuzzy logic methods [25]. In this method, image will be pre-processed, which means it will clear all the errors, duplicate data, extra spaces it will all under the process of sobel filter.

Values of Each Image

Step 1: Get input image
Step 2: Extract the features from the image using image read function
Step 3: The linear kernel function is used to classify the image emotion

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

Step 4: The input image vector is compared with the image vector in database and
Step 5: The Minimal distance between the input & training data is used to identify the emotion
Step 6: Emotion is identified

Table 2: Mean Values Range of Different Emotions

Emotions	Mean Values
Disgust	1-54
Neutral	55-100
Happy	101-139
Sad	140-179
Angry	180-222
Surprised	223-250

In this phase one can see about image reactions and find out what kind of emotions having there in the face using SVM based algorithm, wherein SVM refers to support vector machine [26].

Phase 3- Clustering The Result Using Ts Fuzzy Model

In this section we combine the phase 1 and phase 2 with using a TS fuzzy model.

Takagi-Sugeno Fuzzy Model (TS method)

This model was proposed by Takagi, Sugeno and Kang in 1985. The format of this rule is given as follows:

If x is ‘A’ and y is ‘B’ Then z = f (x,y)

Here, AB are fuzzy sets in antecedents and z = f (x,y) is a crisp function in the consequent.

If 7=x and 9=y then output is z=ax+by+c

Finally, the study showed the emotions through this model indicated below:

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

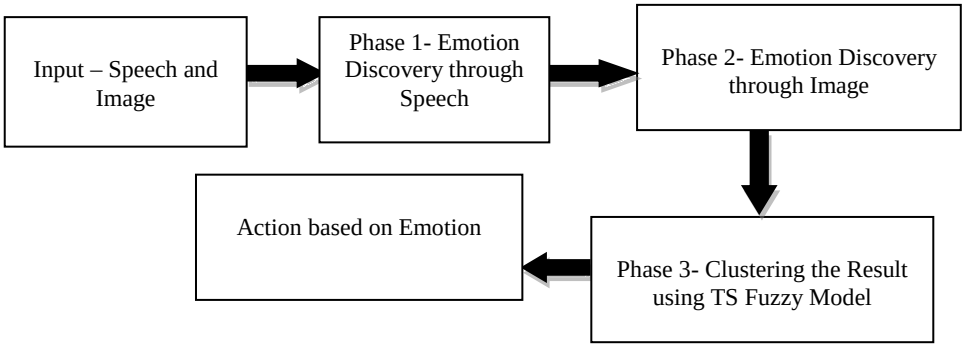

Fig.4 Takagi-Sugeno Fuzzy Model

The combination of Phase I and Phase II is the clustering concept of Phase III.

Takagi-Sugeno Fuzzy Model (TS method)

This model was proposed by Takagi, Sugeno and Kang in 1985. Format of this rule is given as-

If x is ‘A’ and y is ‘B’ Then z = f (x,y)

Here, AB are fuzzy sets in antecedents and z = f (x,y) is a crisp function in the consequent.

Fuzzy Inference Process

The fuzzy inference process under Takagi-Sugeno Fuzzy model(TS Method) works in the following ways:

Step 1: Fuzzifying the inputs- Here, the inputs of the system are made fuzzy.
Step 2: Applying the fuzzy operator- In this step, the fuzzy operators must be applied to get the output.

The rule format of Sugeno form is given as follows:

If 7=x and 9=y then output is z=ax+by+c
Finally, one can find out the emotions through this aforementioned model.

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

Results & Discussion

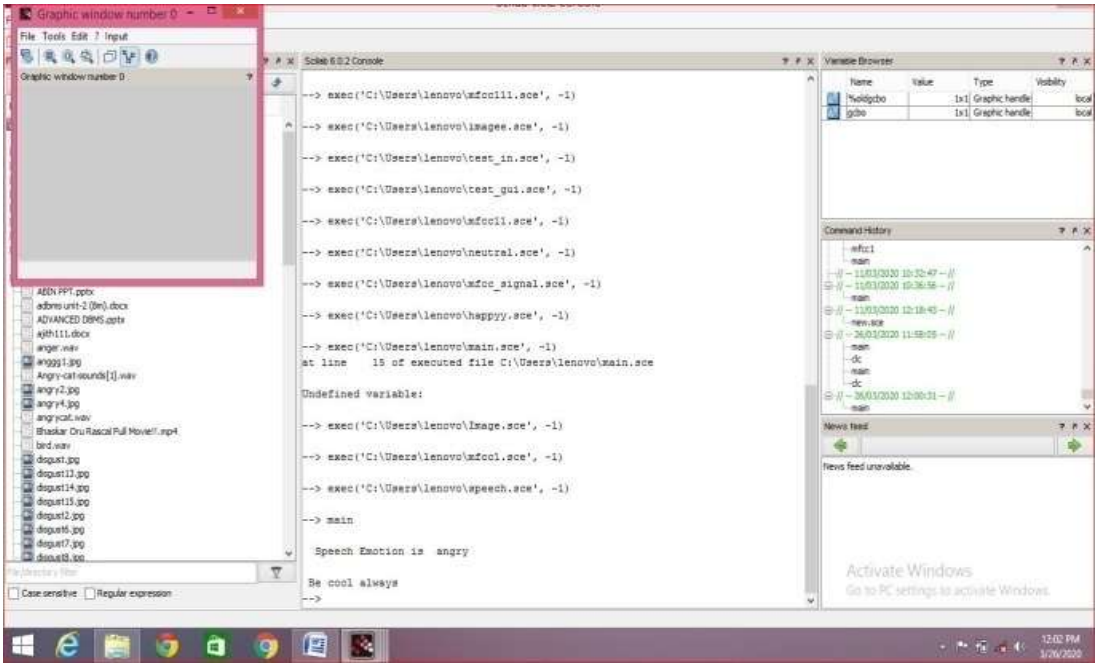

Fig.5 Action based on Speech and Image Based Emotion Recognition using TS Fuzzy Model

The above figure shows the combination of input speech signal and input image processing. In this process it will show the graphic window dialogue box in that window, input menu will be displayed. The speech input is converted into matrix and it is divided into frames, wherein the energy is calculated for each frame. Also, the energy and pitch manipulated for wave signal. The emotion is identified with the pitch range. Moreover, the average energy is manipulated from which pitch is calculated and emotion is recognized. The image will be pre-processed to remove the noise in it. The feature is extracted from image and using Support Vector Machine, the emotion is recognized and Figure 5 shows the result of the action suggested based on emotion identified using speech and image.

Suggestion & Conclusion

The results suggested that proposed system can reliably identify the single emotion from speech samples. The performance highly depends on emotional speech samples. Hence, it is quite necessary to take proper and correct speech samples. The performance has been observed good using the propose technique, but it takes a time in terms of execution. The future researchers shall work on time effectiveness. Another future enhancement is that it can be applied for bigger set of emotions i.e., positive or negative and can be implemented by other classification algorithm. The proposed scheme presented an approach to recognize the emotion from the human

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

speech and face. This approach has been implemented by neural network. This research work focused on the feature extraction method that is useful in the emotion recognition through speech signal, wherein for the purpose of feature extraction, Mel Frequency Cepstrum Coefficient (MFCC) has been used. Hence, to achieve the good extraction of the feature, high pass filter is designed. The high pass filter reduces noise from the signal and helps to extract better feature rather than other filters. Furthermore, to achieve the better performance the neural network is used for training. Moreover, using high pass filter before the feature extraction and neural network for the classification gives the higher accuracy.

References

[1]. Chavhan, Y., Dhore, M. L. & Yesaware, P. (2010). Speech Emotion Recognition Using Support Vector Machine, International Journal of Computer Applications, Vol.1, No.20, pp.345-356.

[2]. Muda, L., Begam, M. & Elamvazuthi, I. (2010). Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, Journal of Computing, Vol.2, No.3, pp.138-143.

[3]. Goldberg, L. (1990). An Alternative Description of Personality: The Big Factor Structure, Personality and Social Psychology, Vol.59, No.6, pp.1216-1229.

[4]. Aly, A. & Tapus, A. (2013). A Model for Synthesizing a Combined Verbal and Nonverbal Behavior based on Personality Traits in Human-Robot Interaction, 8th ACM / IEEE International Conference on Human-Robot Interaction (HRI), Tokyo, pp.278-289.

[5]. Ekman, P., Friesen, W. & Ellsworth, P. (1982) What Emotion Categories or Dimensions can Observers Judge from Facial Behavior? Ekman, P. (1982). Emotion in the Human Face, Ed., Cambridge University Press, New York, pp.39-55.

[6]. Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT), In: Speech Coding and Synthesis, Kleijn, W. B. & Paliwal, K. K. (1995). Eds., Elsevier, pp.495-518.

[7]. Sondhi, M. (1968). New Methods of Pitch Extraction, IEEE Transactions on Audio and Electroacoustics, Vol.16, No.2, pp.262-266.

[8]. Rabiner, L., Atal, B. & Sambur, M. (1977). LPC Prediction Error - Analysis of its Variation with the Position of the Analysis Frame, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.25, No.5, pp.434-442.

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

[9]. Bennacef, S., Bonnea-Maynard, H., Gauvain, J., Lamel, L. & Minker, W. (1994). A Spoken Language System for Information Retrieval, In Proceedings: 3rd International Conference on Spoken Language Processing (ICSLP), Japan, pp.1271-1274.

[10]. Miller, S., Bobrow, R., Schwartz, R., Ingria, R. (1994). Statistical Language Processing Using Hidden Understanding Models, In Proceedings: Human Language Technology Workshop, Plainsboro, New Jersey, USA, pp.278-282.

[11]. Goldberg, E., Driedger, N. & Kittredge, R. (1994) Using Natural Language Processing to Produce Weather Forecasts, IEEE Expert Intelligent Systems and their Applications, Vol.9, No.2, pp.45-53.

[12]. Levin, E. & Pieraccini, R. (1995). Concept-based Spontaneous Speech Understanding System, In Proceedings: 4th European Conference on Speech Communication and Technology (EUROSPEECH), Madrid, Spain, pp.555-558.

[13]. Rong, J., Li, G. & Chen, Y. P. (2008). Acoustic Feature Selection for Automatic Emotion Recognition from Speech, Information Processing and Management, Vol.45, No.3, pp.315-328.

[14]. Busemann, S. (2005). Ten Years After: An update on TG/2, In Proceedings: European Natural Language Generation Workshop, Saarbrucken, Germany, pp.89-96.

[15]. Lavoie, B. & Rambow, O. (1997). A Fast and Portable Realizer for Text Generation, In Proceedings: 5th Conference on Applied Natural Language Processing (ANLP), Stroudsburg, USA, pp.265-268.

[16]. Pathak, A. R., Pandey, M. & Rautaray, S. (2018). Application of Deep Learning for Object Detection, Procedia Computer Science, Vol.132, No.1, pp.1706-1717.

[17]. Pathak, A. R., Pandey, M. & Rautaray, S. (2018). Construing the Big Data based on Taxonomy, Analytics and Approaches, Iran Journal of Computer Science, Vol.1, No.4, pp.237-259.

[18]. Amos, B., Ludwiczuk, B. & Satyanarayanan, M. (2016). Open Face: A General-Purpose Face Recognition Library with Mobile Applications, Technical Report, School of Computer Science, Carnegie Mellon University, Pittsburg, Pennsylvania, pp.1-18.

[19]. Cristianini, N. & Shawe-Taylor, J. (2000). Introduction to Support Vector Machines, Cambridge University Press, United Kingdom, pp.30-35.

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15

[20]. Hsu, R. L., Abdel-Mottaleb, M. & Jain, A. K. (2002). Face Detection in Color Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.5, pp.696-706.

[21]. Alabbasi, H. A. & Moldoveanu, F. (2014). Human Face Detection from Images based on Skin Color, 18th International Conference on System Theory, Control and Computing (ICSTCC), pp.532-537.

[22]. Chen, Z., Liu, C., Chang, F. & Han, X. (2013). Fast Face Detection Algorithm based on Improved Skin-color Model, Arabian Journal for Science and Engineering, Vol.38, No.3, pp.629-635.

[23]. Jiang, H. & Learned-Miller, E. (2017). Face Detection with the Faster R-CNN, 12th IEEE International Conference on Automatic Face & Gesture Recognition, Washington, D.C., USA, pp.650-657.

[24]. Yi, D., Lei, Z., Liao, S. & Li, S. Z. (2014). Learning Face Representation from Scratch, Computer Vision and Pattern Recognition, arXiv:1411.7923, pp.1-9.

[25]. Farfade, S. S., Saberian, M. J. & Li, L.J. (2015). Multi-View Face Detection using Deep Convolutional Neural Networks, In Proceedings: 5th ACM on International Conference on Multimedia Retrieval, Dublin, Ireland, pp.643-650.

[26]. Whiten, A. (2011). The Scope of Culture in Chimpanzees, Humans and Ancestral Apes, Philosophical Transactions of the Royal Society B: Biological Sciences, Vol.366, No.1567, pp.997-1007.

ISBN: 978-81-942938-7-3
DOI: 10.47059/CIIR/BP20002/15