Learning of Unstructured Data Using Machine Learning Algorithm

Tushar Ghorpade[1], Bhavika Tuteja[2], Vaibhav Dholam[3],Gauri Patil[4],Ashutosh Bhujbal [5]

[1]Professor, [2,3,4,5]Students

Department of Computer Engineering,Mumbai University,Navi Mumbai,India.

[1]tushar.ghorpade@gmail.com

[2]tutejabhavika@gmail.com

[3]vakky1232@gmail.com

[4]patilgauri512@gmail.com

[5]theflash257@gmail.com

Abstract— **Unstructured data is data that does not have any pre-defined model associated with it. Audios,images,video files etc can be referred to as unstructured data. Dealing with unstructured data in the form of audios is time consuming and challenging.Crying plays a crucial role in ensuring the survival, health and development of an infant. Therefore,it is important to understand the reason behind the cry. This research develops a system to classify the infant cry sound by performing feature extraction using 13 MFCC(Mel-Frequency Cepstral Coefficients). The coefficients are then passed on to Back Propagation Neural Network with 8 nuerons in the hidden layer, for classification of the audio samples into three classes namely-hungry,anger and fear.**

Keywords— **Unstructured Data, infant cry,feature extraction,Mel-Frequency Cepstral Coefficients, Back Propagation Neural Network.**

I. Introduction

Big data,being heterogeneous in nature,does not circumscribe bounds of data;it comprehends unstructured data which comprises of audios, texts, videos and so forth. Unstructured data is the culmination of myriad forms of information that either does not have a pre-defined data model or is not organized in a predefined manner.This amounts to vacillation and incertitude,which makes it onerous to understand using conventional programs, as opposed to data stored in fielded form in databases or annotated documents.The analysis of the infant cry has been the subject of multitudinous research efforts over the past thirty years. Since the infant cry is the first and the only communication factor with its caregiving environment, it is thought that information pertaining to the state of an infant can be determined by their beseeching wails.Classification of an infant's cry sound is entailed for the amelioration of a parenting or babysitting experience,especially for young parents who have little to no experience in parenting.[1].

Prospects are to develop an automated and non invasive platform to monitor an infant's status to diagnose their needs. Our system will incorporate preprocessing, including frame blocking and windowing followed by MFCC(Mel-Frequency Cepstral Coefficients) feature extraction whose coefficients will be sent to BNN (Backpropagation Neural Network), for classification based on voice type and will classify the infant crying sounds into three classes:hungry,anger and fear.

II. Literature Survey

This section comprises of the findings related to our topic. It is basically an evaluative report of information found in the literature relevant to Classifications of Infant Cry Vocalizations.

The paper [3] proposed several baby cry classification experiments performed on a database of about 40 healthy babies, crying due to five physiological needs: hunger, discomfort, eructation, flatulence and tiredness. The database was created and labeled by our research group based on the Dunstan Baby Language (DBL) babycry classification video tutorial. For automatic classification, two methods that were successful in speaker and language recognition: the GMMUBM (Gaussian Mixture Model - Universal Background Model) and the ivectors modeling methods.

In the paper [4] the convolutional neural networks was adopted to train the infant crying data. Accordingly, the trained CNN is capable to classify the crying into hungry, pain, and sleepy.The crying data was collected from National Taiwan University Hospital Yunlin Branch.The deep learning framework Caffe developed by the Berkeley Vision and Learning Center (BVLC) was used in this paper.Crying was converted to spectrogram.

The paper [5] encompasses modern age feature extraction and machine learning tools for the detection of infant cry signals. In developmental psychology, infant crying is a measure of distress and an automated tool for measurement is extremely important. Initial testing is done using the MIRToolbox in MATLAB which proved to be nonideal for real time signal analysis due to slow speed and poor memory management in MATLAB.

The paper [6] develops a system to identify first whether the cry is an infant cry or not and further classifies the cry into three classes-hungry,tired and discomfort using multiple combinations of Mfcc and Back Propagation Neural Network.
After scrutinizing all the methods mentioned in the literature survey,we present the classification of infant's crying sound in three classes-hunger ,anger and fear We incorporated two methods-Mel-Frequency Cepstral coefficients for feature extraction and Back Propagation Neural Network with 13 input neurons which were the 13 Mfcc Coefficients, 8 nuerons in the hidden layer and 3 output neurons for classification of cries.
The flow of the remaining paper is as follows:First we present the Data Collection,followed by the design,propsosed system, results,conclusion and future work.

III. Data Collection

We garnered a veritable variety of audio samples of infant cries from Donate a cry corpus, repository and a few snippets from youtube videos- all with an average duration of 6 seconds so that the phonetics of the cry sound would be captured easily. The phonetics were understood by the theory of Dunstan Baby Language.We considered 300 audio feature vectors of three classes namely-hungry,anger and fear,out of which 81 audio samples of each class formed the training data and 19 samples formed the testing data.

TABLE I
Data Collection

Class	Training Data	Testing Data
Hungry	81	19
Anger	81	19
Fear	81	19
Total	243	57

IV. Design System

The entire architecture involves three major steps:Preprocessing,Mel-Frequency Cepstral Feature Extraction and Back Propagation Neural Network for Classification. In the first step as in Fig.1, system loads an infant cry sound file. The file type of sound should be .wav (waveform audio format) . The unwanted,silent parts of the sound are filtered and now it is ready for preprocessing. The system will extract MFCC feature. The extraction result gives coefficients.Furthermore, coefficients will be used as input data and trained using BNN as a classifier.

A. Preprocessing

Preprocessing consists of two stages:Frame blocking and Windowing.
1) Frame Blocking:
The unwanted signal or the silent parts of the signal, are discarded.The entirety of the signal,has to be broken down

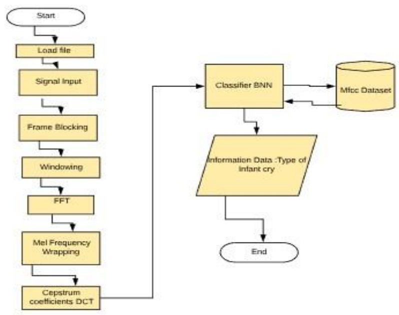

Fig. 1. System Architecture

into frames. As speech is a time varying signal, framing is required since, when examined over a short period of time, the values are stationary.General values are 256 samples, and overlap length is 100 sample.

$$nFrame = floor\left(\left(\frac{(L-N)}{M}\right)+1\right)....[6] \qquad (1)$$

Where,
L = signal data length
N = frame length
M = overlap length

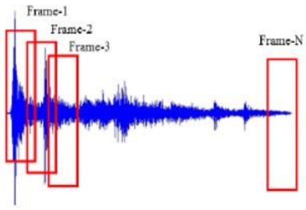

Fig. 2. Frame Blocking[6]

2)Windowing:
The windowing process reduces spectral leakage. Each frame is multiplied by a suitable Hamming window [6].

$$\mathbf{w(n)} = 0.54 - 0.46\cos\left(\frac{2\Pi n}{N}\right), 0 \leq n \leq N-1...[6] \qquad (2)$$

Where , N = frame length n = number of signal sample
In order to reduce the discontinuity of the signal,we apply hamming window function and get the result windowed signal using the formula[6]:

$$x(n) = s(n).w(n) \qquad (3)$$

Where w(n) = hamming window function s(n) = signal sample

B. Mel-Frequency Cepstral Coefficients Feature

The Mfcc feature vectors were used in feature extraction. Each feature vector consisted of 13 cepstral coefficients which acted as input neurons to our Back Propagation Neural Network.
Coefficients of MFCC are obtained by performing FFT (Fast fourier transform) which converts time domain to the frequency domain.FFT is performed in order to obtain the magnitude frequency of each frame.We multiply the response by a set of triangular band pass filters to get a smooth spectrum and then apply DCT(Discrete Cosine transform) to get the coefficients which will then be normalized and would be sent to BNN for classification.

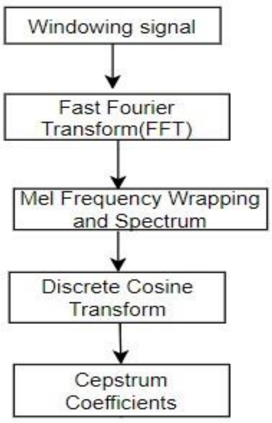

Fig. 3. Mel-Frequency Cepstral Coefficients stages

C.Back Propagation Neural Network

The feature vectors obtained are then given as input to our Back Propagation Neural Network. The Network consists of 13 input neurons,1 hidden layer consisting of 8 neurons and 3 output neurons depicting the three classes of infant cries. The average of the neurons in the input and output layers which (13+3)/2 i.e 8 helped us to decide the number of nuerons in our hidden layer.

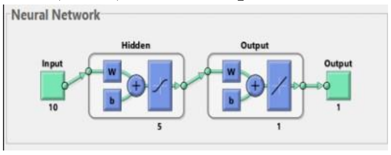

Fig. 4. Back Propagation Neural Network[6]

V. Proposed Methodology

The proposed system is real time system which directly puts the audio which you want to test in the testing data itself, the only restriction being that the audio file should be in a .wav format. The audio sample then goes through preprocessing, and after obtaining the windowed signal, the Mel-Frequency Cepstral coefficients are obtained. We have used 13 Mfcc coefficients in our system. An audio sample consists of various Mfcc feature vectors.

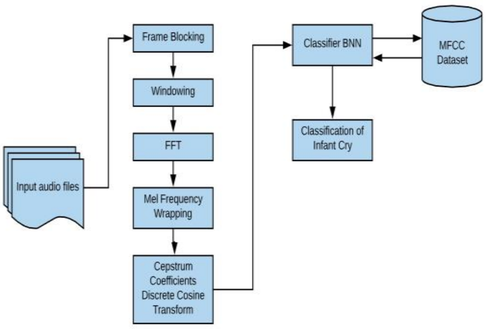

Fig. 5. Proposed Methodology Work Flow

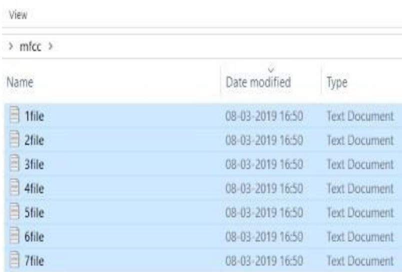

Fig. 6. Mfcc Feature vector generation of an audio sample

A script is run to get the feature vectors containing the coefficients as shown in Fig.6. The feature vectors(training data) obtained are then appended with the class indicator ,0 being for Hungry, 1 for Anger and 2 for Fear. Finally,our Back Propagation algorithm generates the weights which are updated and then testing data is tested for giving us the desired results.

VI. Experimental Results

We were successfully able to classify the infant cry sounds of three classes namely:hungry, anger and fear with and overall accuracy of 77.167%.We were successfully able to obtain 100% accuracy of two classes namely hungry and fear. We obtained a low accuracy of 31.5% for anger as the Mfcc feature vector coefficients were very near to each other's numerical values. The true false values for each class is shown in TABLE II.

TABLE II
PROPOSED SYSTEM ACCURACY

Class	True	False	Total	Accuracy(%)
Hungry	19	0	19	100
Anger	6	13	19	31.5
Fear	19	0	19	100

The result is depicted with one of the solution which can be used as shown in Fig 7.

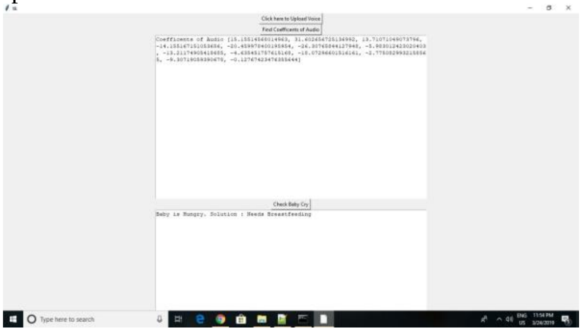

Fig. 7. Result classifying the cry as hungry

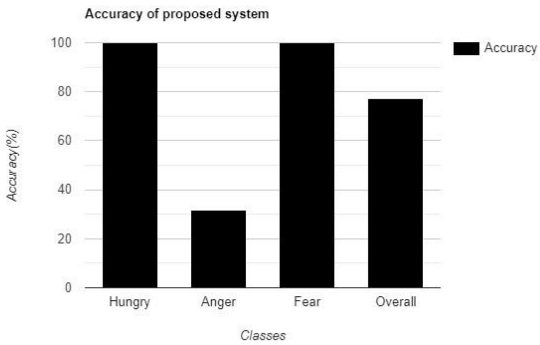

Fig. 8. Overall accuracy of proposed system

VII. DISCUSSION

Based on the results of various experiments, MFCC feature extraction proved to be an efficient method. The classification results obtained are satisfactory but are low for the anger class and work has to be done for improving the accuracy of the anger class. The concomitant of increasing the dataset and the number of MFCC coefficients, and its effect on the system remains to be seen.

VIII. CONCLUSION AND FUTURE WORK

In lieu of the current modus operandi of classifying crying sounds, a methodology which is capable of classifying baby crying sound events, according to the pathological status of the infant for helping pediatricians during the diagnosis process, has been designed. We have presented the development and application of Back Propagation Neural Networks and MFCC for the classification and discrimination between three types of infant cries:hungry,anger and fear. It can be concluded that,the ability system to do classification of infants cry sound is directly proportional to the the number of coefficients of the MFCC. The use of the neurons number in the hidden layer also affects the number of iterations at the training process. An overall accuracy level of 77.167% with 13 coefficients of MFCC neuron input and 8 neurons in the hidden layer is achieved.

Acknowledgment

We take this opportunity to express our profound gratitude and deep regards to our supervisor Mr.Tushar Ghorpade for the exemplary guidance, monitoring,research brainstorming and constant encouragement throughout the completion of the project. We are truly grateful to his efforts to improve my technical writing skills.

References

[1]Unravelling Unstructured Data: A Wealth of Information in Big Data..Amity Institute of Information Technology Amity University Uttar Pradesh, Noida, India ,IEEE. [2]Mukesh Agrawal The infants cry in health and disease...The National Medical Journal of India.

[3]Ioana-AlinaBnic, HoriaCucu, Andi,Buzo, DragoBurileanu and Corneliu Burileanu- -Automatic Methods for Infant Cry Classification...IEEE(2016)

[4]Chuan-Yu Chang, Jia-Jing Li Application of Deep Learning for Recognizing Infant Cries...National Yunlin University of Science Technology, Taiwan,IEEE-2016.

[5] Thesis on Infant Cry Detection...University of Miami,2016

[6]Yesy Diah Rosita,Hartarto Junaedi Infants Cry Sound Classification using Mel-Frequency Cepstrum Coefficients Feature Extraction and Backpropagation Neural Network.

[7]Pritam Pal, Ananth N. Iyer and Robert E. Yantorno Emotion Detection from Infant Facial Expressions And Cries...2015.

[8]M. Petroni, A.S. Malowany, C.C. Johnston2, B.J. Stevens3 Classification of Infant cry Vocalizations using Artificial Neural Networks (ANNs).. Department of Electrical Engineering and Center for Intelligent Machines (CIM), McGill University. [9]Xuan Zhou,Jian Wang,Hongzhi Hu,Weihui Dai.S. Malowany, C.C. Johnston2, B.J. Stevens3 Recognition of Infants Emotions and Needs from Speech Signals.

[10]Audio Pattern Recognition of Baby Crying soundevents...June,2015

[11]Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques.