IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-4 E-ISSN: 2347-2693

Analysis of SMO and BPNN Model for Speech Emotion Recognition System

Rohit katyal

University Institute of Engineering and Technology Department,
Chandigarh University, India

Available online at: www.ijcseonline.org

Received: Apr/02/2016 Revised: Apr /10/2016 Accepted: Apr/22/2016 Published: Apr/30/2016

***Abstract*—** Speech emotion detection refers to discovering the speech category based on the training and testing to the database provided. This research work has been classified in four sections namely SAD, HAPPY, FEAR and AGGRESSIVE. There are two major sections in this research work namely Training and Testing. The training has been done on the basis of wave files provided for every group. Features have been extracted for all groups and have been saved into the database. The testing section classifies the training set of data with the help of BACK PROPAGATION NEURAL NETWORK (BPN) classifier and SEQUENTIAL MINIMAL OPTIMIZATION (SMO) classifier. The results of the BACK PROPAGATION NEURAL NETWORK CLASSIFIER have been found superior in terms of classification accuracy.

***Keywords*—** Speech; Features; SMO; BPNN; Accuracy.

I. INTRODUCTION

Now a day's recognition of emotion with speech is gritting attention from engineers in the field of speech signal processing and pattern recognition. As computers have become an important part of our life so there is a necessity of communication between humans and computer. With the help of voice signals –automatic emotion recognition recognizes the speaker's emotional state. Emotions play a Vitol role in human's life. Humans express their mental state or feelings through emotions. Humans possess a natural gift of sensing emotions through speech while machines fail to do so as they don't have basic intelligence for monitoring emotions. Comprehending emotions through speech is a complicated task as there isn't any obvious solution for what the right emotion would be for a given speech sample. By speech recognition and speaker identification techniques machines can comprehend what's said but if system of emotion recognition is used then emotion can comprehend how it is said. In humans for various actions emotions play an important role so there's a requirement of a system interface where there can be good decision making and human machine communication .definition of emotion recognition through speech consist of disclosure of human's emotional state via feature extraction through her or his voice signal. Few other applications that require this interface are – interactive storytelling, movie and electronic machine pet, E-tutoring application and remote tech school. Other applications where it's been used are - psychiatric diagnosis, car board system, lie detection, aircraft cockpits call centre and intelligent toys. Various intelligent systems have been made on the grounds of different emotions like- happiness, surprise, anger, disgust, stressed, fearful, neutral etc [1]. Emotion recognition can be through spectral features and speech signal prosodic features as these contain a lot of emotional feature like – energy, pitch, loudness, fundamental frequency, glottal parameters and speech intensity these are prosodic features. Few phonetic and lingual features are also used foe emotion detection through speech. Different types of classifiers are utilized for emotion recognition like – [KNN] k-nearest neighbors, [HMM] hidden mark over model, GMM super vector based SVM classifier, and support vector machine [SVM] and [GMM] Gaussian mixtures model [2].

Speech being a voice sample can be used for data analysis. Recording of speech is done either through voice recorder or software. As processing of signals is done mostly in digital form so this can also considered as digital signal processing applied to speech signal. Speech processing consists of manipulation, acquisition, transfer, output and storage of digital speech signals.

Emotional speech recognition recognizes involuntarily physical or emotional condition of a being through her or his voice. In the speech of a speaker there are different stages that are identified as emotional aspects of speech and are unified in Para linguistic aspects. The linguistic content cannot modify by emotional state; in communication of individual this is a significant factor, since feedback information is provided in numerous applications. Speech is possibly the generally proficient way to correspond with each other [3]. This too means that speech could be a helpful boundary to cooperate with machines. A few victorious examples based on it throughout the past years, while we have awareness about electromagnetism; includes the development of the megaphone, telephone. Even in the previous centuries people were researching on speech fusion. Von Kempelen developed an engine talented of 'speaking' words and phrases. At the present time, it has happened to achievable not only to expand examination and execute speech recognition systems, but also to have systems

competent to real-time alteration of text into speech. Regrettably, in spite of the high-quality development made on that area, there are countless applications that are the speech recognition procedure facing dig now; speech is a very prejudiced experience that is added by the majority of them. It is very complex task to recognize the words from speech. In this system first we have to change the representation of speech signal. First task is to convert the speech file into phonemes or syllable then used the dictionary database to match speech input to phonemes or syllable. This dictionary database is used to train the system. If the phonemes are matched then corresponding word is recognized.

II. RELATED WORK

Dimitrios Ververidis and Constantine Kotropoulos [4] gave a description on three goals that hit strike our mind when think about emotional speech recognition. Our first job is to collect data and update record where collection of emotional speech data is available. Record contains data about states of emotions, number of speakers, speech kind etc. In step two, symbolization of goal step is done which are used for extracting features for emotional speech recognition and also to measure how emotion have affect on them. In market, the features that are there are – vocal tract cross section areas, pitch, speech rate and intensity of speech signal. an appropriate algorithm is used in the last goal which will classify speech into emotional states. Various classification methods are examined where timing information is exploited. Classification techniques basis - artificial neural method [ANN], [HMM] hidden mark over models, support vector machines [SVM], k nearest neighbours are reviewed.

SMO or sequential minimal optimization is new techniques proposed by John C Platt [5] for training support vector machines. [Requirement of a solution for the problem of training a support vector machine is to be used on a very large quadratic programming (QP) optimization problem] initially large QP problems are divided into series of small QP problems. Critical examination of the solutions of these problems is done, where QP optimization a time consuming numerical is avoided as an inner loop. Memory requirement for SMO is linear for training set, where large training sets are allowed to be handled by SMO .due to avoidance of matrix computation, for different test problems, in training set size SMO scales lie somewhere between quadratic and linear .SMOs computation for time is represented by SVM evaluation; hence SMOs are faster as compared to SVMs and also bare data sets. [In real-world sparse data sets, it is possible that, SMO could be 1000 times faster when compared with chunking algorithm.]

Wouter Gevaert, Georgi Tsenov, Valeri Mladenov [6] described in this paper an investigation that is done on performance for classification of speech recognition. There are two standard neural networks structures that are used for performance evaluation as classifiers. Feed-forward Neural Network (FFNN) type is included for standard utilization with back propagation algorithm and Basis Functions Neural Networks is Redial used.

Mirza Cilimkovic [7] presented method for classification and clustering in data mining. Neural Networks (NN) as a classifier is used. The proposed system is capable of mimic brain activities and is able to learn. Learning of NN is made from examples. If more examples are provided to NN, then it has capability to knob those examples and classifies that data with representation of patterns in data. There are three layers in basic NN that are as input, output and hidden layer. There are numerous nodes existing in each layer and nodes of input layer need to be attached with nodes from hidden layer. Then to obtain output there should be connections between nodes of hidden layer to nodes from output layer. Weights between these nodes will show the connections. Here algorithm of NN is described, Back Propagation (BP) Algorithm. The main purpose is to show logic behind this algorithm. The basic idea behind BP algorithm is quite simple; output of NN is evaluated against desired output. If we do not get results as expected then there is requirement of modification in connection weights and process is repeated again and again till the error value is small. Some parameters are changed to improve results for new implementation.

Lei-Ming Hong, Qian-Li Ma, Peng Peng, [8] put forth a novel method [for solving method of Support Vector Machine algorithm i.e. SMO that is a parallel algorithm].BY this algorithm master CPU dispenses primitive training sets to slave CPUs. These slave CPUs perform series of SMO on appropriate training sets. As shrink and buffer methods are also selected, increase in speed of parallel training algorithm is performed, which are presented in outcomes of parallel SMO depending on MNIST data set. Results proved that using SMO performance is good for solving large scale SVM.

Rong-En Fan, Pai-Hsuen Chen, Chih-Jen Lin [9] presented a new algorithm for selection of working set in SMO type decomposition method. It discussed that in training support vector machines (SVMs), selection of working set in decomposition process is important. Fast convergence is achieved by using information of second order. Theoretical properties such as linear convergence are established. It is proved in results that proposed method provided better results in contrast to existing selection methods using first order information.

Xigao Shao, KunWu, and Bifeng Liao [10] proposed an algorithm for selection of working set in SMO-type decomposition. It showed that in training part, least square support vector machines (LS-SVMs) the selection of working

set in decomposition process is important. In the proposed method a single direction is selected to achieve the convergence of the optimality condition Experimental results represented that speed of training is faster than others but classification accuracy is not better than existing ones, it's almost same with others.

Gert R. G. Lanckriet, Michael I. Jordan ,Francis R. Bach [11] presented kernel matrices combination for SVM and this combination decrease quadratic ally strained quadratic program n (QCQP) which is a convex optimization problem. Whereas classifiers that are classical kernel based depends on single kernel .mainly combination of multiple kernels are used to develop base of classifier. Unfortunately, small number kernels problem can be solved with data points and current convex optimization toolboxes; because of cost function [SMO] techniques cannot be used which are important in large scale implementations of SVM. [A novel dual formulation of the QCQP as a second-order is proposed for cone programming problem, and shows how to exploit the technique of Moreau-Yosida regularization to succumb a formulation to which SMO techniques can be applied]. For interior points that are present in current optimization toolboxes, Algorithm that is SMO based is more efficient and better compared to general purpose methods.

C. Bhattacharyya S. K. Shevade, K. R. K. Murthy and S. S. Keerthi, [12] in this paper Schölkopf's and Smola [SMO] sequential minimal optimization algorithm have few sources of inefficiency that is pointed out for regression of [SVM] support vector machine which occurs by the use of a single threshold value. For dual problem KKT conditions is used, on the basis of two threshold parameters SMO modification is done which are employed for regression. Compared to original SMO this proposed algorithm with SMO modification performs faster.

III. SMO AND BPNN BASED MODEL

Speech recognition process is basically done by the Speech Recognition System. In the speech recognition process, speech input signal is processed into recognition of speech as a text form. Speech Recognition System helps the technology to bring computers and humans more closely. There is basic terminology that one must know in order to implement or develop a Speech Recognition System [15].

- Utterances- User input speech is called utterances, in simple words when user speaks something it is called utterances.
- Pronunciations- Single word has multiple meanings and multiple recognitions. It all depends on pronunciation. A single word is uttered in different means in accordance to country, age etc.
- Accuracy- It is the performance measurement tool. It is measured by number of means but in this case, if speaker utters "NO", then Speech Recognition System must recognise it as word "NO". If it is done precisely then accuracy of system is efficiently very good or else.

A. *Sequential Minimization Algorithm (SMO)*

Speech Recognition System consists of training part and testing part [16]. So, basically SMO is used for training of SVM part. Primarily SMO is designed to speed up the rate of algorithm based on optimization techniques. The SMO consists of two B constraints B_1 and B_2 and give optimize values for both Bs. This task is reviewed until Bs full exposure. Process of SMO is described below:

- Selecting B constraint
- Optimize B_1 and B_2
- Find threshold value T

B. *Selecting B Constrained*

SMO has optimization techniques that led to enhance the speed of algorithm. So there are two constraints B_1 and B_2, choose that constraint that will lead to achieve the result as far as possible. To maximize the optimization function, is vital step to enhance the speed of computer. As there are n (n-1) choices of possible solutions so there is minute up gradation [17]. SMO tries to optimize the Bs but if no B_1 and B_2 are changed then terminates the SMO.

C. *Optimize B1 and B2 Constraints*

B_1 and B_2 optimization is done using Lagrange operators.

Then restrict these constraints. Restrictions on the constraints are shown as:

If $x^{(1)} /= x^{(2)}$, $L = \max(0, B_1 - B_2)$,
$H = \min(C, C + B_1 - B_2)$

If $x^{(1)} = x^{(2)}$, $L = \max(0, B_1 + B_2 - C)$,
$H = \min(C, B_1 + B_2)$

Now, find the B_2, to maximize the optimization of objective function. If this rate finishes up lying exterior the boundaries L and H, is range for B_2 within which it exists, optimal B_2 value is given as:
$B_2 := B_2 - x^{(2)} (E_1 - E_2) / \eta$

D. *Compute the Threshold T*

After optimizing B_1 and B_2, we select thresholding T. If after optimization B_1 is not at the bounds (i.e., 0< B, < C) then following threshold T_1 is valid.

$T_1 = T - E_1 - x^{(1)} (B_1 - B^{(OLD)} (y^{(1)} - x^{(2)} (B_2 - B_1^{(OLD)} (Y^{(1)}, Y(^2))$

These thresholds are valid in which one is $0 < B_1 < C$ and second one is $0 < B_2 < C$, and they will be identical [18].

E. Back Propagation Algorithm

Back propagation neural network (BNN) is the for the most accepted and commonly used neural network[2-7] .Usually back-propagation neural network consists of feed forward multi-layer network in an input layer, an output layer, and at least one hidden layer. All layers are completely associated to the successive layer [19]. There is no interconnection or feedback between PEs in the same level. Each PE has a predisposition input with non-zero weight. The bias input is analogous to the. In a neural network consisting of P processing elements, the input/output function is defined as:

$$k = L\ (mW)$$

Where m = ***{m,)*** is the input vector to the network, k = {Kkj} is the output vector from the network, W is the weight matrix. It is afterward defined as

$$W = (w_i^T, w_j^T, \ldots\ldots, W_n^T)^T$$

Where the vectors $w_i, w_j, \ldots\ldots, W_n$ are the individual PE weight vectors, which are given as:

$$W_r = \begin{cases} w_i \\ w_j \\ w_n \end{cases}$$

The main function of back propagation neural network is that it moves from input to output layer. In the back propagation neural network error is found at output layer, so to remove error at output layer , so to remove error at output layer, moves backward takes place , hence name is back propagation neural network [20].

IV. PERFORMANCE ANALYSIS

These two algorithms have been tested on 50 speech signals in which training is done with 10 speech signals. In this firstly sample is selected and then features are extracted like Min. Frequency, Max Frequency, Avg. Frequency, Time, Noise, Accuracy, Amplitude and other more features. Fig 1 shows these extracted features of one selected sample.

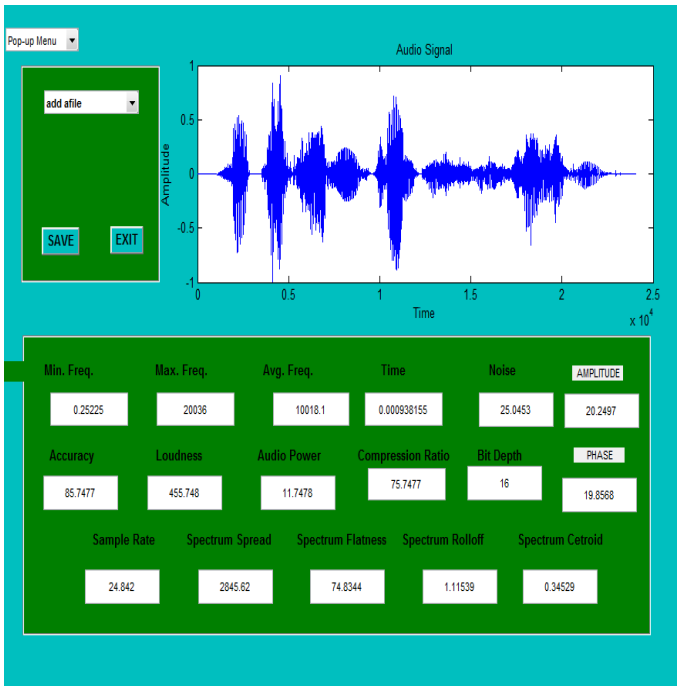


Fig 1: Feature Extraction

There are four major sections in testing phase. The first section is the uploading of the wave file whose features has to be extracted as shown in fig 1and then has to be passed to both the classifiers to check the accuracy of the processing. Fig 2 shows the selected testing sample.

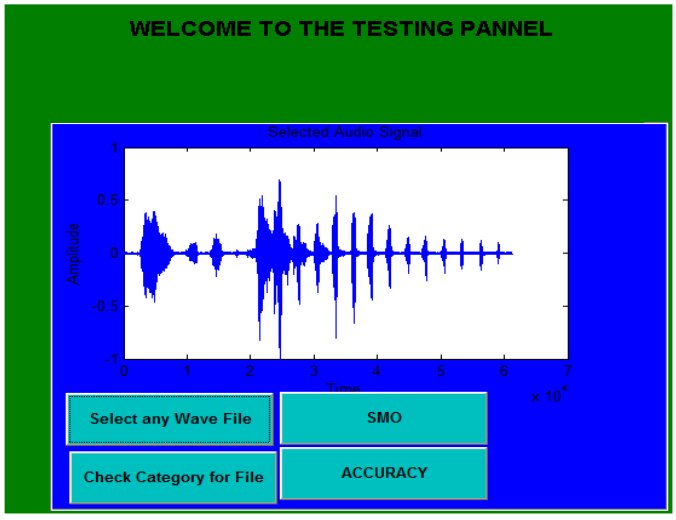
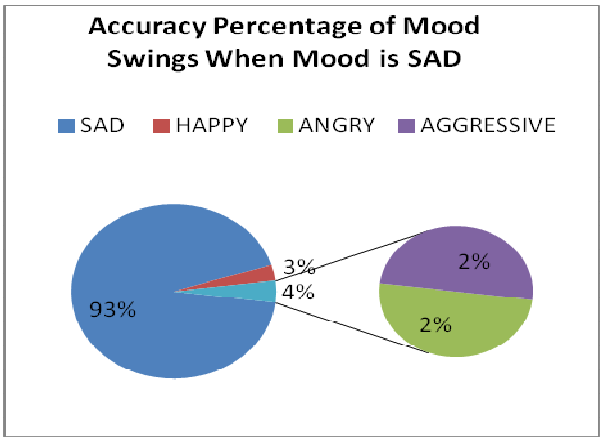
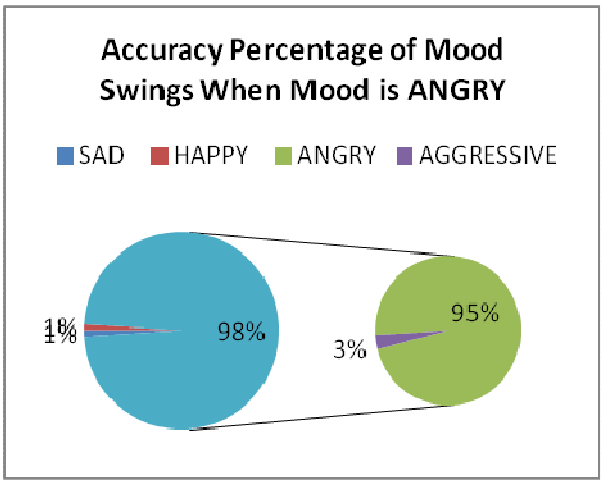


Fig 2: Testing Sample

Fig 3 shows the calculated accuracy of the BPNN and SMO algorithm. The graph explains itself that the accuracy of





BPNN is more than that of SMO for the same classification and it differs from 10 to 15 percent.

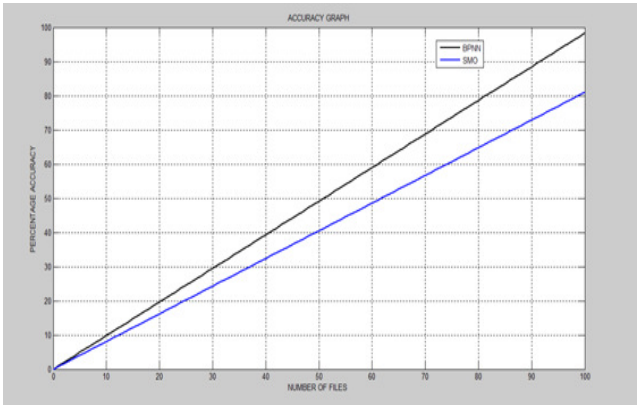
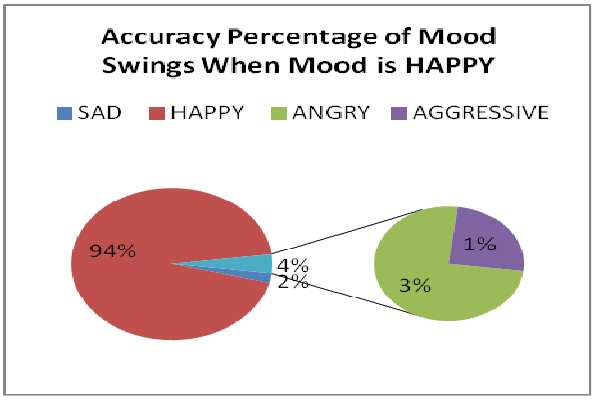


Fig 3: Accuracy of BPNN and SMO

The experiment has been done on the basis of four speech categories namely sad, happy, angry and aggressive in MATLAB 2010. Fig 4 shows accuracy percentage of mood swings like SAD, HAPPY, ANGRY, AGGRESIVE.



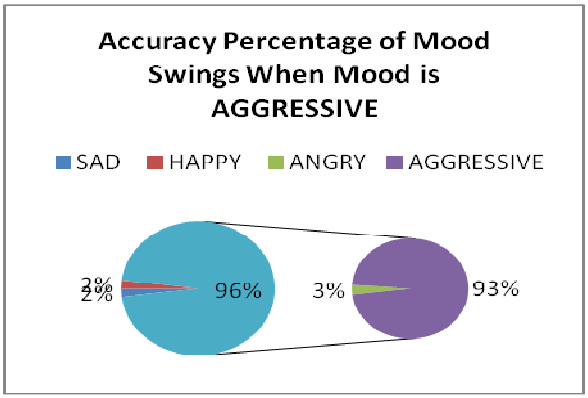


Fig 4: Accuracy Percentage of Mood Swings using BPNN

CONCLUSION

This paper concludes that speech emotion detection by using BPA (Back Propagation Algorithm) that belongs to neural networks perform better with high accuracy. The classification has been done using two algorithms. The accuracy of the system varies for BPNN or BPA and SMO. The classification accuracy for SMO has come out to be 75-83 percent where as the accuracy for BPA algorithm varies from 85 to 95 percent. This research work is limited for the users that mean the current system has to be trained for the every user you want to identify. The other aspect of this research work is that it does not provide any gender information in terms of the classified data set that means it does not specify that the classified voice is for male or for female. For this purpose, the data base has to be trained with a huge data set. In the similar manner, no segmentation has been performed in the analysis of this work to identify the spoken words. The future research workers can also try their hands on combining the values of BPNN and SVM.

REFERENCES

[1] D. A. Sauter; F. Eisner, A. J. Calder and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion," *The Quarterly Journal of Experimental Psychology*, Vol. 63 (11), pp. 2251–2272.

[2] A. Utane, S.L Nalbalwar, "Emotion Recognition Through Speech Using Gaussian Mixture Model And Hidden Markov Model", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.
[3] J. Bachorowski, "Vocal Expression and Perception of Emotion," *Current Directions in Psychological Science,* Vol. 8 (2), pp. 53–57, 1999.
[4] Ververidis and C. Kotropoulos "Emotional speech recognition: Resources, features,and methods," *Speech Communication* , Vol 48, pp. 1162-1181.
[5] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", *Advances in Neural Information Processing Systems* 11, M. S. Kearns, S. A. Solla, D. A. Cohn, eds., MIT Press, (1999).
[6] W. Gevaert, G. Tsenov and V. Mladenov, "Neural Networks used for Speech Recognition," *Journal of Automatic Control,* Vol. 20, pp. 1-7, 2010.
[7] M. Cilimkovic, "Neural Networks and Back Propagation Algorithm", Institute of Technology Blanchardstown, Blanchardstown Road North Dublin 15, Ireland.
[8] P. Peng, Q. L. Ma and L. Hong, "The Research Of The Parallel Smo Algorithm For Solving Svm," *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding*, 2009.
[9] R. Fan, P. Chen and C. Lin, "Working Set Selection Using Second Order Information for Training Support Vector Machines," *Journal of Machine Learning Research* Vol. 6 , pp. 1889–1918, 2005.
[10] X. Shao, KunWu, and B. Liao, "Single Directional SMO Algorithm for Least Squares Support Vector Machines," *Computational Intelligence and Neuroscience* Vol., 2013, Article ID 968438.
[11] F. R. Bach, G. R. G. Lanckriet and M. I. Jordan, "Multiple Kernel Learning, Conic Duality, and the SMO Algorithm", *Proceedings of the 21st International Conference on Machine Learning*, 2004.
[12] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO Algorithm for SVM Regression," *IEEE Transactions on Neural Networks*, Vol. 11, pp. 1188-1193 , 2000.
[13] Y. S. Rao and i. Patel, "Speech Recognition Using Hmm With Mfcc- An Analysis Using Frequency Specral Decomposion Technique," *Signal & Image Processing : An International Journal(SIPIJ)* Vol.1(2), 2010.
[14] W. HAN, C. CHAN, C. CHOY and Kong-Pang PUN, "An Efficient MFCC Extraction Method in Speech Recognition," *International Symposium on Circuits and Systems Proceedings*, 2006.
[15] H. Gupta and D. S. Wadhwa, "Speech Feature Extraction and Recognition Using Genetic Algorithm," *International Journal of Emerging Technology and Advanced Engineering*, Vol. 4 (1), 2014.
[16] J. C. Platt, "Fast Training of Support Vector Machine using Sequential Minimal Optimization," *In Advances in Kernel Methods: Support Vector Learning*,, pp. 185-208, 1999.
[17] T. Glasmachers and C. Igel, "Second Order SMO Improves SVM Online and Active Learning," Journal Neural Computation, Vol. 20 (2) pp. 374-382, 2008.
[18] Y. Pan, P. Shen , and Liping Shen," Speech Emotion Recognition Using Support Vector Machine" , International Journal of Smart Home Vol. 6, No. 2, April, 2012.
[19] R. Rojas," Neural Networks: A Systematic Introduction," *Springer Berlin Heidelberg New York*, 2005.
[20] N.Pushpa, R.Revathi, C.Ramya and S.Shahul Hameed, "Speech Processing Of Tamil Language with Back Propagation Neural Network and Semi-Supervised Training," *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.2 (1), 2014..
[21] Divesh N. Agrawal and Deepak Kapgate, "Face Recognition Using PCA Technique", *International Journal of Computer Sciences and Engineering*, Volume-02, Issue-10, Page No (59-61), Oct -2014

AUTHORS PROFILE

Rohit Katyal, A.P., Chandigarh University, B.Tech in Computer Science Engineering, Master of Engineering in Information Security with 2 years of experience in teaching.