IJRECE VOL. 4 ISSUE 2 APR.-JUNE 2016 ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE)
DOI: 10.13140/RG.2.1.2138.0724

A Review Process and basic architecture of Speech Emotion Recognition based on signal Processing

Ajay Kumar[1], Er.Deepika Chaudhary[2], Er. Viney Dhawan[3]
[1]*Student (M.Tech), Department of Electronics and Communication, KITM, Kunjpura, Karnal*
[2]*Assistant Professor, Department of Electronics and Communication, KITM, Kunjpura, Karnal*
[3]*Head of Department, Department of Electronics and Communication, KITM, Kunjpura, Karnal*

Abstract **-** An increasing attention has been directed to the study of the emotional content of speech signals, and hence, many organisations have been projected to recognize the emotional content of a spoken utterance. This paper is a survey of speech emotion organisation addressing three central features of the design of a speech emotion recognition system. In human mechanism boundary application, emotion recognition from the speech signal has been research topic since many years. To recognize the reactions from the speech signal, many systems have been developed. In this paper speech emotion recognition grounded on the preceding technologies which uses different classifiers for the emotion recognition is reviewed. We describe a technique which allows to continuously controlling both the age of a synthetic voice and the quantity of emotions that are communicated. Also, we existing the first large-scale data mining experiment about the automatic recognition of basic emotions in familiar average short words. We focus on the speaker-dependent problem. An optimal feature set is derivative through the use of an inherited algorithm. Finally, we explain how this study can be applied to real world situations in which very insufficient instances are accessible. Furthermore, we describe a game to play with a personal robot which facilitates instruction of instances of emotional sounds in a natural and rather unconstrained manner.

Keywords **-** Speech signals, emotional signals, speech emotion classification and Speaker-dependent problem**.**

I. INTRODUCTION

Humans have the natural ability to use all their accessible senses for supreme mindfulness of the received communication. Through all the accessible senses people really sense the emotional [1] state of their communication partner. The demonstrative detection is normal for humans but it is very problematic task for mechanism. Therefore the determination of emotion gratitude system is to use feeling related information in such a way that human mechanism communication will be improved. Emotion recognition from the speaker's speech is very tough because of the succeeding motives: In separating between various emotions which certain speech topographies are more convenient is not clear. Because of the presence of the different judgments, speakers, dialogue styles, language rates accosting capriciousness was announced, since of which speech structures get directly affected. The same sound may show dissimilar emotions. Each emotion may correspond to the different portions of the spoken[2]sound. Therefore it is very problematic to distinguish these portions of utterance. Another problem is that emotion appearance is contingent on the chatterer and his or her culture and environment. As the philosophy and atmosphere gets change the talking style also gets change, which is extra challenge in opposite of the speech emotion gratitude system. There may be two or additional types of emotions, extended term emotion and passing one, so it is not pure which type of emotion the recognizer will detect.

Human machine interfaces are frequently used currently in many submissions. Most of them necessitate the uncovering of emotion in the speech. But same insufficient human machine interfaces being implemented currently are able to achieve that. Therefore, there is a need to size a human machine boundary that can detect emotions successfully and efficiently. Verification of emotions can be done using three factors [3] the content of the speech, facemask expressions of the lecturer or by the structures extracted from the emotional speech. This paper is confined to the gratitude of emotion by manufacture use of the features extracted from the speech.

Usually human beings can easily detect numerous kinds of emotions. This can be attained by the human mind through years of preparation and thought. The human concentration detentions all kinds of emotions since youthful and is taught to separate between the emotions based on its comments. For occurrence, when a creature is angry, his tone raises, his countenance becomes stern and the content of his speech no longer remains pleasant. Similarly, when an individual is happy, he speaks in a musical tone, there is an expression of happiness on his face and the content of his dialogue is rather pleasant and festive. Based on these observations, a person can speedily recognize the state of the talker whether he is happy, sad, angry, depressed, disgusted etc. A human machine boundary that can process dialogue having emotional content makes use of a similar concept [4]training and then testing. In the exercise phase, the interface is fed with models of each emotion. The classifier used in the boundary extracts features from all the models and forms a mixture for each emotion. In the testing phase, demonstrative speech is given

DOI: 10.13140/RG.2.1.2138.0724

as input to the classifier. The classifier quotations the features from the input and compares it to all the combinations. The input is confidential into that emotion to which it is closest. In other words, the input file will be confidential into that emotion whose topographies are the most similar to that of the input file. There are a number of structures and classifiers that can be cast-off for the purpose of emotion detection. However, it is difficult to identify the best model between these since the collection of the feature set and the classifier depends on the problem.

II. RELATED WORK

MirzaCilimkovic [5] 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 specimens. If more samples are providing 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. **PengPeng, Qian-Li Ma, Lei-Ming Hong, (2009) [6]** presented a novel technique for solving method of Support Vector Machine algorithm that is SMO that is a parallel algorithm. According to this algorithm, primitive training setsare dispensed by master CPU to slave CPUs. Slave CPU run serial SMO on the relevant training sets. As buffer andshrink methods are also selected, increment in speed of the parallel training algorithm is done, which is represented in the results of parallel SMO based onthe dataset of MNIST. The results of this work proved that by using SMO performance of solving large scale SVM is good. **Rong-En Fan, Pai-Hsuen Chen, Chih-Jen Lin, (2005) [7]** presented a new algorithm for selection of working set in SMO type decomposition method. It conversed that in exercise supportvector 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 firstorder information.**Xigao Shao, KunWu, and Bifeng Liao, (2013) [8]** proposed an algorithm for selection of working set in SMO-type decomposition. It showed that in training part, least square supportvector 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. **Francis R. Bach, Gert R. G. Lanckriet, Michael I. Jordan, (2004) [9]**showed combination of kernel matrices for SVM and that combination reduces a convex optimization problem known as quadratic ally strained quadratic program n (QCQP). While classical kernel-based classifiers arebased on a single kernel, mostly base of classifier are developed by a combination of multiple kernels. Unfortunately, the problem for small number of kernels can be solved with current curving optimizationtoolboxes and a small number of data points; besides, due to cost function thesequential minimal optimization (SMO) systems which are indispensable in large-scale implementationsof the SVM cannot be applied. A novel dual formulation ofthe QCQP as a second-order is proposed for cone programmingproblem, and shows how to exploit the technique of Moreau Yolinda regularization to succumb a formulation to which SMO techniquescan be applied. SMO- based algorithm is much better and efficient than general purpose methods for interior point's availablein current optimization toolboxes.**S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, (2000) [10]**proposed Shola and Schölkopf's sequential minimal optimization(SMO) algorithm have some source of inefficiency which is pointed out for regression of support vector machine (SVM) that occurs by the use of a single threshold value. The KKT conditions for the dual problem is used, SMO modification is done on the basis of two threshold parameters that are employed for regression. This proposed algorithm with the modification in SMO performs faster than the original SMO.

III. GOAL

It is necessary that robotic pets can also recognize the emotions articulated by the humans who are networking with them. Human beings generally use all the context and modalities, from lexica to facial expression and intonation. Unfortunately, this is not an easy thing for a machine in an uncontrolled atmosphere: for occurrence evigorous speech recognition in such situations is out of reach for current systems. Facial expression recognition requirements both computational resources and video devices those robotic [7] creatures most often do not have. For this motive, we explored how far we could go by using only prosodic information in the voice. Furthermore, the speech we are attentive in is the kind that happens in everyday conversations, which means short informal utterances, as contrasting to the speech fashioned when one is asked to read a paragraph with emotions from a newspaper. Four broad classes of emotional satisfied were intentional: joy/pleasure, sorrow/sadness/grief, anger and calm/ neutral.

IV. EMOTIONS

An emotion is simply a feeling or impression caused by a person's acuity about something or someone. Emotions are our thoughts felt physically. Emotions are liveliness. Emotions can convert into other emotions.

Causes of Feelings

DOI: 10.13140/RG.2.1.2138.0724

Emotions are simply energy in motion**.** When you sensation an feeling it is because that energy has been triggered by a thought and it sends it in motion, which creates feelings felt in various[8] locations in your body. Emotions can travel. They can travel from one part of the body to another. They can become headaches, stomach-aches, pain in other parts of the body, or wonderful tingly sensations throughout your body.

Types of Emotions

a) Happiness and Sadness

There are only two basic emotions. One is happiness the other is sadness**.** There is pretty a widespread array of these emotions and what usually causes our feelings to change is into what time period we are jutting them. For example, sadness echoed into the future is knowledgeable as fear. On the other hand, reflecting grief into the ancient is experienced as anger. These feelings can be experienced in other ways too depending on which we are bulging them toward. For sample, when we point anger inward, anger becomes guilt [9].

b) Positive and Negative Emotions

Positive and negative emotions cannot exist within you at the same time. You cannot feel pain and pleasure at the particular same time. You can nevertheless, feel ache one moment and pleasure the very next. You can switch back and forth production you think that it is practised concurrently. However, it's not. You can experience an array of negative feelings all at the same time, or an array of confident emotions.

V. FEATURES FOR SPEECH EMOTION RECOGNITION

An important question in the enterprise of a speech emotion recognition system is the extraction of suitable features that efficiently illustrate different emotions. Subsequently pattern recognition techniques are rarely independent of the problem domain, it is whispered that a proper collection of features significantly affects the classification performance. Four issues must be measured in feature extraction. The first question is the region of analysis used for feature extraction. While some scientists follow the commonplace framework of dividing the speech signal into small intervals, called frames, from each which anindigenous feature trajectory is extracted, other researchers prefer to extract global statics from the whole speech utterance. Another imperativeenquiry is what the best article types for this task are, e.g. pitch, energy, zero crossing, etc.? A third inquiry is what is the consequence of ordinary dialogue processing such as post-filtering and silence removal on the overall presentation of the classifier? Lastly, whether it suffices to use acoustic features for modelling emotions or if it is necessary [10] to combine them with other types of features such as linguistic, discourse information, or facial features.

VI. ARCHITECTURE OF SPEECH EMOTION RECOGNITION

The block diagram of the emotion recognition system through speech measured in this study is demonstrated in Figure. 1. The block diagram consists of the emotional speech as effort, feature extraction, article selection, classifier and detection of emotion as the output.

1) Emotional Speech Input

A suitable emotional speech database is significant requirement for any emotional recognition model. The quality of database regulates the effectiveness of the system. The emotional database may contain collection of acted speech or real data world.

Fig.1 Architecture of speech emotion recognition

2) Feature Extraction and Selection

An imperative step in emotion acknowledgment system through speech is to select a significant feature which carries large demonstrative information around the speech signal. After collection [11] of the database containing emotional speech appropriate and necessary structures such as prosodic and spectral features are extracted from the speech signal. The commonly used structures are pitch, energy, MFCC, LPCC, formant. The steps involved in calculation of MFCC are shown below.

3) Classification

The most imperative characteristic of emotion recognition system through speech is classification of an emotion. The performance of organization is reliant on on proper choice of classifier. There are many types of classifier such as Hidden Markov Classical (HMM), Gaussian Mixture Classical (GMM), Artificial Neural Network (ANN) and Support Vector Machine (SVM) [12].

VII. CONCLUSION

In this paper, a survey of current research work in speech emotion recognition system has been given. Speech emotion appreciation systems grounded on the numerous classifiers is showed. 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. 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

DOI: 10.13140/RG.2.1.2138.0724

extracting more effective features of speech, accuracy of the speech emotion recognition system can be enhanced.

VIII. REFERENCES

[1]. Ibrahim Patel,Dr. Y. SrinivasRao, "Speech Recognition Using Hmm With Mfcc- AnAnalysis Using Frequency SpecralDecomposion Technique", Signal & Image Processing : An International Journal(SIPIJ) Vol.1, No.2, December 2010.

[2]. Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong-Pang PUN, "An Efficient MFCC Extraction Method in SpeechRecognition", IEEE, 2006.

[3]. Dorigo, Marco, Mauro Birattari, and Thomas Stützle. "Ant colony optimization." Computational Intelligence Magazine, IEEE 1.4 (2006): 28-39.

[4]. Che, Zhen-Guo, Tzu-An Chiang, and Zhen-HuaChe. "Feed-forward neural networks training: A comparison between genetic algorithm and back-propagation learning algorithm." International Journal of Innovative Computing, Information and Control 7.10 (2011): 5839-5850.

[5]. MirzaCilimkovic, "Neural Networks and Back Propagation Algorithm", Institute of Technology Blanchardstown, Blanchardstown Road NorthDublin 15, Ireland.

[6]. PengPeng, Qian-Li Ma, Lei-Ming Hong, "The Research Of The Parallel Smo Algorithm For Solving Svm", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.

[7]. Rong-En Fan, Pai-HsuenChen,Chih-Jen Lin, "Working Set Selection Using Second Order Informationfor Training Support Vector Machines", Journal of Machine Learning Research 6 , pp.1889–1918, 2005.

[8]. Xigao Shao, KunWu, and Bifeng Liao, "Single Directional SMO Algorithm for Least Squares SupportVector Machines", Computational Intelligence and Neuroscience, Article ID 968438. 2013.

[9]. Francis R. Bach, Gert R. G. Lanckriet, Michael I. Jordan, "Multiple Kernel Learning, Conic Duality, and the SMO Algorithm", Proceedings of the 21st International Conferenceon Machine Learning, Banff, Canada, 2004.

[10]. 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, No. 5, September 2000.

[11]. DimitriosVerveridis and Constantine Kotropoulos, "Emotional speech recognition: Resources, features, and methods", Artificial Intelligence and Information Analysis Laboratory, Department of Informatics, Aristotle University ofThessaloniki, Box 451, Thessaloniki 541 24, Greece.

[12]. WouterGevaert, GeorgiTsenov, ValeriMladenov, "Neural Networks used for Speech Recognition", JOURNAL OF AUTOMATIC CONTROL, UNIVERSITY OF BELGRADE, VOL. 20:1-7, 2010.