MACHINE LEARNING BASED GENDER RECOGNITION AND EMOTION DETECTION

Damanpreet Kaur
Research Scholar, Computer Science and Engineering,
Chandigarh University, Gharuan, India

ABSTRACT

The voice of a person plays an important role in analyzing people. From voice of the person we not only recognize the gender of the person but also detect the emotion of the person. Gender recognition and emotion detection both has its importance in forensics, games, in security purposes and of course in our day to day life. In our work we are using fundamental frequency as feature for gender recognition. We have taken MLP as classifier and achieved 92.5% accuracy. Furthermore for emotion detection pitch, speaking rate and energy are taken as features and adaboost with C4.5 ML algorithm is taken for classification and achieved 93.3% accuracy. Moreover, we have analyzed the relationship between gender and the emotion of the persons. We analyzed that emotional accuracy for females is lower as compared to the males i.e. emotion of the male can be detected more accurately than the female from voice.

KEYWORDS: *Gender Recognition, emotion detection, adaboost, C4.5.*

I. INTRODUCTION

Gender Recognition (GR) means recognizing the gender of the person whether the person is men or women. It is a significant task for human beings, as many communal functions precariously rely upon correct gender awareness. Automatic human GR by machines has currently received symbolic attention in computer vision community. Features like face, 3D body shape, gait (manner of walking), footwear, fundamental frequency of voice etc are used for gender recognition.

The ability to do automatic recognition of human gender is essential for a number of systems that process or exploit human-source information. Common examples are information retrieval, human-computer or human-robot intercommunication. The result of an AGR system can be used for achieving meta-data information useful for annotating audio files. Moreover, gender is an important cue that can be exploited for improving intelligibility of man-machine interaction, or just, for decreasing the investigation space in applications such as speaker recognition or surveillance systems. Gender recognition has great importance in forensics, games, business intelligence, demographic survey, visual surveillance

It is completely known that humans speech enclose linguistic content, identity as well as the emotion of speaker. Emotion plays a significant role in daily interpersonal human interactions. This is essential to our rational as well as intelligent decisions. It helps us to match and understand the feelings of others by conveying our feelings and giving feedback to others. And speech is relevant communicational channel enriched with emotions. The tone, timing, energy, speaking rate etc jointly express the different emotions. For example, if a person is in the condition of fear, joy or anger then the voice produced becomes louder, faster precisely propound with a wider as well as higher pitch range. There are many applications of detecting the emotion of the persons like in audio surveillance, web-based E-learning, commercial applications, clinical studies, entertainment etc. Emotion identification can be used as voice tag in different database access systems. This voice tag is used in telephony shopping, ATM machine as a password for accessing that particular account.

Initial methods of gender recognition include extracting features like eyes, hair, lips, chin, face etc; moreover from footwear too we can recognize the gender of the person. In recent trends voices are used to recognize the gender and emotion. We aim to analyze the role of fundamental frequency and other features in identification of the gender of a person. And the features such as pitch, energy and speaking rate are extracted for detecting the emotion of the person.

Rest of the paper is organized as: section 2, discusses the literature survey. Section 3 introduces the proposed methodology. In section 4 work done has been discussed in which brief description of speech signal acquisition, pre-processing, feature extraction and classification is given. Experimental results are given in section 5 in which results are shown in graphs. Section 6 contains conclusion and future work.

II. LITERATURE REVIEW

Automatic gender recognition of a speaker has various potential applications. Gender dependent models are more precise than gender independent models in the framework of automatic speech recognition. The aim of AGR is to determine the gender of person from the given speech/audio signal. The result of AGR system could be beneficial for many applications, such as making gender particular sound models for automatic voice recognition, decreasing the research space in surveillance systems or speaker recognition, examining human- computer intercommunication, and the social interaction and behavior. The earliest computer-vision techniques of gender recognition were neural networks based. Golomb et al. in [1] worked on facial images and build a fully connected network SEXNET having two layers in it for gender recognition. Tamura et al. in [2] used face images of different resolutions and worked on multilayered neural n/w to recognize gender. Gutta et al. in [3] projected a hybrid approach made up of neural n/w & decision trees both. Huchuan Lu et al. in [4] proposed PPBTF for the real-time GR. A pattern map is built from gray scale image, where for the texture information and for characterizing, edges and lines are used. The image basis functions were obtained by PCA and then used as templates for the pattern matching. They used Adaboost algorithm to find the best discriminative feature subdivision (subset), and for classification support vector machine (SVMs) was used. After experimentation on the images taken from the front from FERET dataset it shows that PPBTF is really effective and faster in computation as compared to Gabor. Pronobis et al. in [5] analyzed fundamental frequency (Fo) and the cepstral features for the robust GR. They carry out the experiments on BANACA corpus datasets. Jian-Gang et al. [6] Proposed score level fusion with Adaboost for speaker gender recognition. They demonstrated that this method could perform better than other methods which used the single information on voice. Yi-jie Zhao et al. [7] proposed a scheme having classification rate more than 80% based on color Information for gender recognition. Charles et al. [8] showed that age of the person can also affect the accuracy of the gender recognition. They study the database having 8,000 images containing the age group of 0 to 93 years. Wei et al. [9] proposed a lip movement gender recognition method to improve accuracy by exploring the dynamic information while a user is speaking. There are many other kinds of gender classifiers that has been proposed. For example, Wu et al. [10] had selected a series of LUT weak classifiers by employing adaboost technique for gender categorization. The Adaboost-classifier used was trained on over 11,000 face images, and obtained a recognition rate of 88% for 200 LUTs. Jain and Huang et al. [11] used linear discriminant and independent component analysis to develop a gender classifiers. For testing, 500 face images having 64 × 96 size were used from the FERET database, an eye detection system was used to keep constant separation between two eyes by aligning the face images correctly. 99.3% accuracy was reported at end. Moghaddam and Yang et al. [12] investigated the support vector machine (SVM) classifier by using face images for gender recognition. The face images having size 21×12 and 1755 in number were used to evaluate SVM-classifiers. The rate of classification was 95.12% obtained with the cubic polynomial kernel and with Gaussian RBF kernel, the rate obtained was 96.62%. Fok Hing Chi Tivive et al. [13] proposed an AGR system that can use the images of random size and can detect the faces in them, and detect their gender. Two modules of the system are: face detector and gender classifiers. First, these networks use same network architecture to do feature extraction and classification. Second, introducing variations in two dimensional shapes and a certain degree of changes to distortions taken place. Moreover, the stage of preprocessing lies between face detection and GR is wiped out. Mehmet et al. [14] analyzed the perceptual audio features for detecting the emotion of the person. They proposed some acoustic features for recognizing the emotion from audio. Carlos et al. [15] analyzed the emotionally salient visible features of fundamental frequency for emotion detection. David Philippou-hubner et al. [16] analyzed the work or importance of speaking rate in Emotion recognition from speech signal, as the speaking rate is an important prosodic feature of speech and it is easy for humans to estimate how fast a person is talking. Et al. [17] focused on 2-way classification. They showed that features derived from agitated emotions have similar properties and they worked on pitch, MFCC and formants. Dipti D. Joshi et al. [18] proposed a review of speech emotion recognition and discussed the most recent work done in the field of emotion detection from speech and reviewed the different methods of feature extraction and classification. Rathina et al. [19] analyzed the prosodic features in emotional speech. Prosodic features include pitch contour, utterance timing and energy contour.

As per the literature survey, it was found that females usually have shorter and thinner vocal cords than males. And the fundamental frequency (F_0) of female voices is typically higher than fundamental frequency (F_0) of male voices. This makes fundamental frequency (F_0) a prospective choice for gender recognition. And hence we choose the fundamental frequency as the feature to extract. And for emotion detection pitch, energy and speaking rate are the main features to extract from the audio signal. As by getting the pitch, speaking rate and energy of the voice of the person one can detect the emotion of the person easily For example, speech produced in a state of fear, anger or joy becomes faster, louder, precisely enunciated with a higher and wider pitch range.

III. PROPOSED METHODOLOGY

Gender recognition and emotion detection consists of four main steps. First, the voice sample collection, second pre-processing to enhance the quality of the voice samples. Then features vector is formed by extracting the features which is input to machine learning classifier for recognition. This process is described below with the help of a block diagram shown in Fig.1

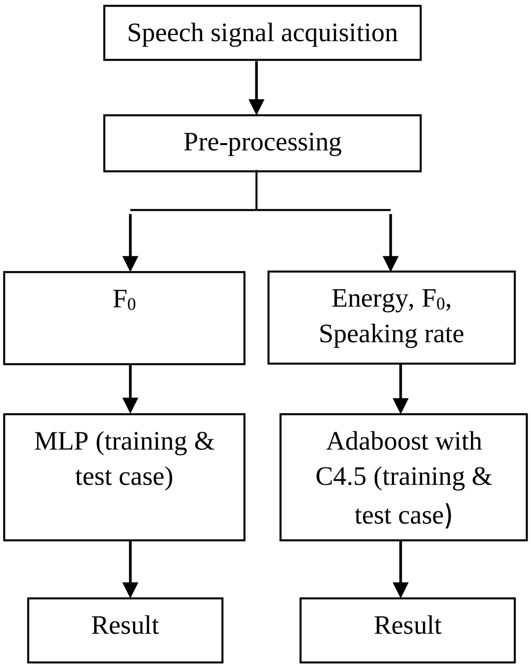
Speech signal acquisition
Pre-processing
F0
Energy, F0, Speaking rate
MLP (training & test case)
Adaboost with C4.5 (training & test case)
Result
Result

Fig: 1. Proposed method for Gender recognition and emotion detection.

IV. WORK DONE

a. Speech signal acquisition**:** Speech signals were captured from 800 people including both male and female in the age group of 20-30 years with the help of microphones inbuilt in Acer PC. All the voice samples whether for gender recognition or emotion detection were recorded under normal circumstances that means samples may contains the noise like of fan or any other common noise. For gender recognition people were asked to speak ‘hello’ word and for emotion detection people were told to speak in three different emotions ‘happy’, ‘normal’ and ‘sad’. The sampling is done while recording the voice sample as well. Sampling is described as the formation of discrete signal from the continuous signal. The speech signals captured were sampled at 44 kHz. Praat has been used for recording the audio signals.

b. Preprocessing**:** Preprocessing has been done for removing the noise present in the collected samples. This includes pre emphasis and windowing. Pre emphasis process removes the noise from the captured signal. Spectral subtraction method has been used to remove the noise. Further hamming window is used for windowing.

c. Feature extraction**:** F_0 is extracted for gender recognition and pitch, energy and speaking rate are extracted for emotion detection. Features were extracted from the voice signals using praat tool [20].

d. Classification**:** Feature vector is constructed using the above mentioned features which were fed to ML algorithms for recognition. MLP is used for gender recognition and Adaboost with C4.5 is used for emotion detection. Classification is done by using the weka [21] which was developed at New Zealand based University of Waikato. It is an open simulator and has a java based implementation of ML algorithms and researchers use it extensively.

V. EXPERIMENTAL RESULTS

The experiment was carried out by taking voice samples from 800 male and female speakers of age group 20-30 years. Voice samples were taken under normal circumstances for both gender and emotion detection. Speakers were told to speak ‘hello’ word for gender recognition and for emotion detection they were asked to speak in three different emotions ‘happy’, ‘normal’ and ‘sad’ mood. Praat tool has been used to collect the voice samples. From voice samples fundamental frequency was extracted as feature for gender recognition and pitch, energy and speaking rate were extracted for emotion detection. For gender recognition machine learning algorithm named as MLP was used for classification and adaboost with C4.5 was used for emotion detection.

It was found that recognition rate for gender recognition using MLP is 92.5%. Further for emotion detection MLP do not perform well as its recognition rate was low i.e. 55.8% as MLP performs well if number of input units will be less but in emotion detection we have more input units. Then we used the C4.5 algorithm as it is fast in performing classification task and got 78.15% recognition rate. To enhance the performance of C4.5 we used the adaboost algorithm as it is a boosting algorithm that can be used to improve the performance and we got the very good results with average accuracy of 93.13%. Comparison is shown in figure below:

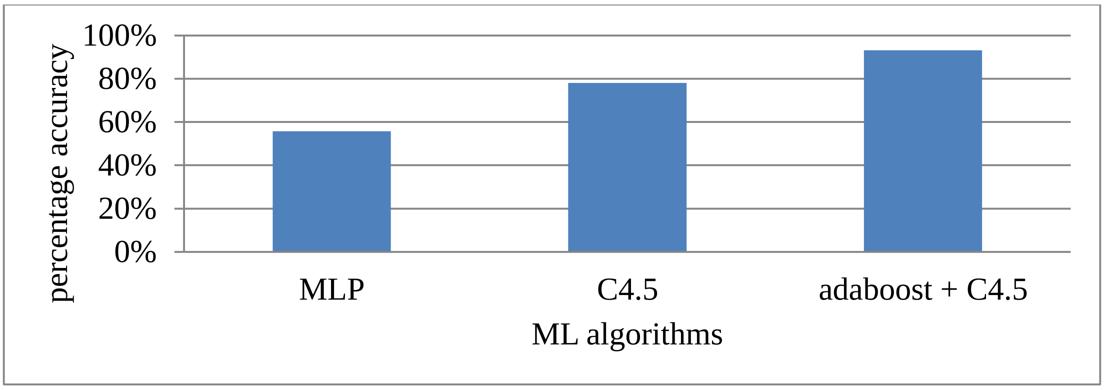

Fig 2 Comparison of ML algorithms for emotion detection

The percentage of the emotional accuracy for both male and female shown in figure 3 below:

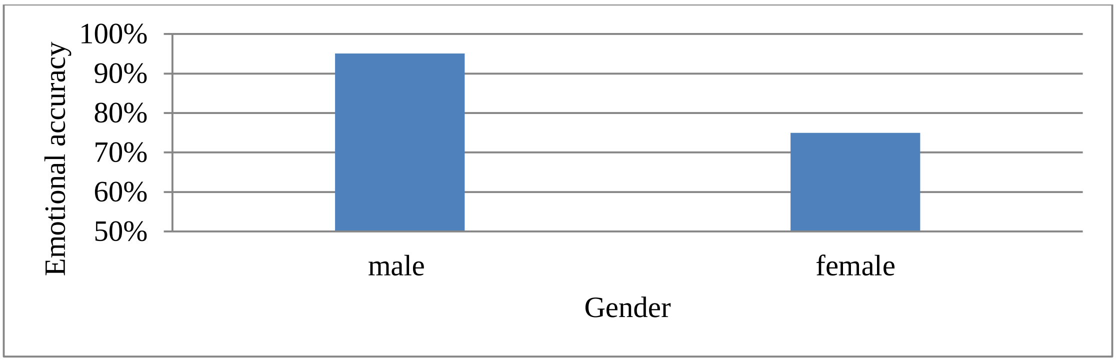

Fig.3 Emotional accuracy of both male and female

As we analyzed emotional accuracy varies with gender so this bar chat shows that accuracy in the case of males is higher than the females. Its difficult to detect the emotion of the females as compared to males as sometimes females express two different emotions in same manner, they react exactly same at two different moments like they speak loudly in happy mood but also in sad mood. Moreover sometimes they speak slowly in angry mood too and in sad mood. While in case of males they mostly react differently or we can say express different emotions in different manners. That's why it becomes easy to detect the emotion of the males.

VI. CONCLUSION

The work presented in our paper proposed a feature extraction and recognition system to recognize the gender and emotion of the persons. The features that are taken in our proposed work comprise fundamental frequency for gender recognition and pitch, energy and speaking rate for emotion detection. For classification MLP is used for gender recognition and adaboost with C4.5 is used for emotion detection. Algorithms were used and got 92.5% average accuracy for gander recognition and 93.13% for emotion detection. Moreover it is also concluded that emotion varies with gender, from experimentation, it is difficult to detection emotion of females as compared to males.

VII. FUTURE WORK

This work can be expended further in future by increasing the dataset as we have taken the voice samples of 800 people, one can take the more number of voice samples. Further samples can be taken in the vacuum, as the noise affects the accuracy of the gender recognition and emotion detection, it may give the better results. Moreover different feature set can be included as we have taken pitch, energy and speaking rate for emotion detection and fundamental frequency ($F_{o)}$ for gender recognition, in future more features can be added like MFCC coefficients, LPC coefficients etc. This work can be further expended by using the other emotions like anger, boredom etc as we have worked on only three emotions, happy, normal and sad.

REFERENCES

[1]. Golomb B., Lawrence D., Sejnowski T.: Sexnet: a neural network identifies sex from human faces. In: Advances in Neural Infor- mation Processing Systems, pp. 572–577. Morgan Kaufmann, San Mateo (1991)

[2]. Tamura S.H., Kawai, Mitsumoto H.: Male/female identification from 8 9 6 very low resolution face images by neural network. Pattern Recognition. 29, 331–335 (1996)

[3]. Gutta, S., Weschler, H., Phillips, P.J.: Gender and ethnic classification of human faces using hybrid classifiers. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 194–199 (1998)

[4]. Huchuan Lu, Yingjie Huang, Yen wei Chen, Deli Yang "automatic gender recognition based on pixel patteren based texture feature", Springer-Verlag 2008.

[5]. Marianna Pronobis, Mathew Magimai.-Doss., "Analysis of F0 and cepstral features for robust automatic gender recognition."

[6]. Masatsugu Ichino, Naohisa Komatsu, Wang Jian-gang,and Yau Wei Yun., " Speaker Gender Recognition Using Score Level Fusion By AdaBoost," ,11th Int. Conf. Control, Automation, Robotics And Vision Singapore., 2010, pp.648-653.

[7]. Guo Shiang Lin, Yi Jie Zhao., "A Feature Based Gender Recognition Method Based On Color Information," First International Conference On Robot, Vision And Signal Processing, 2011, pp. 40-43.

[8]. Guodong Guo , Charles R. Dyer , Yun Fu, Thomas S. Huang., " Is Gender Recognition Affected By Age," IEEE 12th International Conference On Computer Vision Workshops, ICCV Workshops., 2009, pp. 2032-2039.

[9]. Masatsugu Ichino, Naohisa Komatsu, Wang Jian-gang,and Yau Wei Yun., " Text Independent Speaker Gender Recognition Using Lip Movement," ,IEEE 12th Int. Conf. Control, Automation, Robotics And Vision Singapore., 2012, pp. 176-181.

[10]. Wu B., Ai H., Huang C, "LUT-based Adaboost for gender classification", AVBPA 2688, 104–110 (2003).

[11]. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Conf. on Comput. Vision and Pattern Recognit., 2005, pages 886–893.

[12]. Moghaddam B., Yang M.H.: Gender classification with support vector machines. In: IEEE Trans. PAMI 24, 707–711 (2002).

[13]. Fok Hing Chi Tivivive, and Abdesselam Bouzerdoum, "gender recognition system using shunting inhibitory convolutional neural networks."

[14]. Mehmet Cenk Sezgin, Bilge Gunsel* and Gunes Karabulut Kurt, "Perceptual audio features for emotion detection", EURASIP Journal on Audio, Speech, and Music Processing, 2012.

[15]. carlos busso,sungbok lee shrikanth narayanan,"analysis of emotionally sailent aspects of fundamental frequency for emotion detection", IEEE transansactions on audio, speech, and language processiong, vol. 17, no.4, may 2009.

[16]. David Philippou-H¨ubner, Bogdan Vlasenko, Ronald B¨ock, Andreas Wendemuth," The Performance of the Speaking Rate Parameter in Emotion Recognition from Speech", IEEE International Conference on Multimedia and Expo Workshops, pp. 296-301, 2012.

[17]. Emotion detection from speech, URL:http://cs229.stanford.edu/proj2007/ShahHewlett%20%20Emotion%20Detection%20from%20Speech.pdf.

[18]. Dipti D. Joshi1, Prof. M. B. Zalte," Speech Emotion Recognition: A Review", IOSR Journal of Electronics and Communication Engineering, Volume 4, Issue 4 (Jan. - Feb. 2013), PP 34-37,2012.

[19]. X. rathina, K.M. mehata, M. Ponnavaikko," basic analysis on prosodic features in emotional speech" , international journal of computer science, engineering and applications vol.2, No. 4, august 2012, PP 99-107,2012.

[20]. Praat URL:http://www.fon.hum.uva.nl/praat/download_win.html.

[21]. WEKA 3: Data Mining With Open Source Machine Learning Software in JAVA.