International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-6 Issue-6, August 2017

A Survey Paper on Gender Identification System using Speech Signal

Mohit Kumar Mishra, Arun Kumar Shukla

Abstract: ***Gender is a critical statistic characteristic of individuals. This paper provides a survey of automatic human gender identification using speech signal characteristics and classifiers. A review of approaches exploiting information from human speech presented. Here, highlights of selection of speech features, their processing and different classifiers used for this purpose are discussed. Based on the results discussed in the papers it can be stated as, accuracy of automatic gender identification system with any classifiers is better if speech dataset used for training and testing is taken/ recorded in the same environments. Pitch is the basic feature of speech which distinguishes between adult man and woman. Other features like MFCC, LPC, RASTA-PLP also used for automatic gender identification. Neural Network, Support Vector Machine (SVM), Random Forest etc. are used for automatic gender identification through speech signal. Till now, many challenges are still available here to identify gender with acceptable accuracy in real life environmental speech where noise is acoustically added with human speech.***

Index Terms: Gender identification, MFCC, SVM.

I. INTRODUCTION

Gender identification, language identification, emotion recognition etc. of speech through trained classifiers using speech signal properties has been increased attention in recent years. Accurate classification of above said attributes can be used for other applications based on speech signals such as human-computer interaction, surveillance, speech recognition, speaker recognition, speech coding, language dependent searching, biometrics, gender specific advertising etc. For example, in automatic speaker recognition system, search space can be reduced to half furthermore; it will reduce the total search time. It can also be used for automating advertisements based on gender using speech signal properties. Human can easily identify the gender of persons by listening their speech but it is a tough task for automated Machine. Here, a brief survey of human gender identification using speech signal is presented. Basic differences between male and female speakers are their pitch and frequencies. So, researchers in early stage used pitch and formant frequencies features of speech to distinguish between the male speaker and female speaker.

Manuscript published on 30 August 2017.
* Correspondence Author (s)
Mohit Kumar Mishra, Department of Computer Science & Information Technology, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Naini, Allahabad (U.P)-211007, India. E-mail: mk83541985@gmail.com
Arun Kumar Shukla, Department of Computer Science & Information Technology, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Naini, Allahabad (U.P)-211007, India. E-mail: aks.jit@gmail.com

Since, these features are not robust in case of noise and unvoiced sound then different cepstral domain features are also used for the said purpose. Few researchers used combination of time domain feature like Pitch and cepstral domain features like MFCC, different frequencies etc. and fuse these features together to form feature vector.

Now, these feature vectors are used for training and testing different types of classifiers like neural network, support vector machine, random forest etc. Few researchers also tried to calculate minimum duration of speech required for identify the gender of a speaker with highest accuracy. In section II, Survey for gender identification is performed followed by conclusion in section III.

II. RESEARCH SURVEY

In [1], Ming Li et al. used seven different classification approaches separately and also by fusing all these models for automatic gender identification based on speech. They used acoustic features of speech along with prosodic level features. They used SVM-Prosody, GMM, SVM-UWPP-SVM, GMM-Mean-SVM, GMM-UWPP-Sparse representation and GMM-MLLR-SVM. Best result was obtained when all models were combined as claimed in the Paper.

H. Meinedo et al. [2] fused short term prosodic and acoustic features along with long term features. Training of four different models is performed using the said features for gender and age classification of speaker. Trained models were tested for different datasets and claimed that best result is obtained using linear logistic regression classifier.

Since, using these features and the models gender identification for adult and child is very tough because pitch range is almost similar for male child or female child. Nisimura and A. Lee worked on a system that can be guided using speech. This system is also able to distinguish in adult and child [3]. Here, combination of linguistic features and acoustic features are fused to create feature vectors and Support Vector Machine classifier was used for the classification purpose. 92.4 % discrimination accuracy was claimed in the said paper. Some researchers works for gender and age classification system for the robotic applications which provide services to complete daily home tasks. In 2007, H. Kim and his colleagues used an age and gender classification system for a home robot service. GMM, ANN and MLP used as clasiifier and for training of these systems MFCC feature of speech was used [4]. They claimed that accuracy of gender identification was 94.9 % with GMM classifier. 81.09% accuracy was achieved with Jitter and Shimmer features of speech by ANN classifier. Wei Li and his colleagues [5]

worked for automatic gender recognition task along with language identification task. Acoustic features of speech were used for the training and testing purpose while GMM is used for the classification purpose.

Phuoc Nguyen and his colleagues [6] worked for age, gender and accent recognition system. They fused GMM and SVM classifier for the proposed work and vector quantization is used. Australian speech database was used for training and testing of this system. Australian speech database contains 108 speakers and each speaker has 200 utterances. 97.96 % to 98.68 % accuracy was claimed by them in the proposed system.

In 2011, Weighted Pair wise Principle Component Analysis (WPPCA) speech dimension reduction method was proposed by Gil Dobry and his colleagues [7]. SVM with RBF kernel was used for the classification. They worked on age classification system. They claimed that due to reduction of feature vector size training of the classifier became fast and over-fitting also reduced.

M. H. Bahari and his colleague proposed an approach for age estimation and gender identification system using hybridization of Weighted Supervised Non-Negative Matrix Factorization (WSNMF) and General Regression Neural Network (GRNN). Dutch database was used for the feature extraction, training and testing of the classifier. 96% accuracy is obtained for gender detection as claimed by the author in this Paper [8].

In 2009, M. H. Sedaaghi et. al. used probabilistic neural network, Support Vector Machine, KNN and GMM classifier for recognizing the age and gender of a speaker based on speech [9]. They used DES and ELSDSR [22] database for training, validation and testing of the system. They found that Support Vector Machine and probabilistic neural network provides better result for both age and gender recognition system among above mentioned classifier in the system.

In 2010, Tobias Bocklet et al. worked on automatic gender and age classification system using speech and used many different classifiers individually as well as their combination for the classification purpose [10]. Glottal features, prosodic features and spectral features of speech extracted and used for the training and testing purpose of the classifier. GMM-UBM combination of the classifier provides best result for age and gender detection as claimed by them in this paper. They claimed that classification accuracy for this system was 42.4%.

Michael Feld et al. used GMM/SVM combination for automatic speaker gender and age detection in the car [11].

Florian Metze et al. worked for gender and age detection which is to be used in telephone applications. They also made a comparison between humans and the trained systems for the same data for age identification of a person [12]. Following features along with following classifiers were used for this task:

i. Parallel phone recognizer.
ii. A dynamic Bayesian network used combination of prosodic features of speech.
iii. Linear prediction analysis approach.
iv. GMM classifier used with MFCC features.

Parallel phone recognizer recognizes similar to humans being for long time utterances while accuracy decreases when duration of speech decreased as claimed by the authors in the Paper.

Christian Muller et al. proposed age and gender recognition system that can be used for fulfilling the special needs for elder persons [13]. ScanSoft and M3I corpus are used for this task. To generate feature vector Jitter / Shimmer and speech rate were used. ANN, KNN, SVM etc. used as classifier for the classification task and four classes (elder male, non-elder male, elder female and non-elder female) were used.

In 2012, Myung-Won Lee et al. worked for gender and age group recognition system which can be used for for human-robot interaction [14]. SVM and Decision Tree classifiers are used for the classification task which is trained and tested with MFCC and LPCC features of speech.

In [15], Frank Wittig and Christian Muller proposed a gender based age recognition system by combining different classifiers with dynamic Bayesian network.

Tobias Bocklet et al. proposed a system for classification of age to preschool and primary school age children. GMM super-vector is used and traineing is performed either by SVM or SV regression [16].

M. H. Bahari et al. proposed a method for speaker age estimation using HMM and WSNMF [17]. Least Squares Support Vector Repressor (LS-SVR) classifier is used for the classification process.

Seema Khanum et al. [18] proposed for gender identification using speech in a noisy environment. Artificial Neural Network is used for classification process. To train and test the classifier MFCC feature vectors were used.

Samiksha Sharma et al. [19] developed speaker & gender identification system with continuous speech signal for different languages. For classification purpose Radial basis network function is used while MFCCs and delta-MFCCs are used to train and test the trained model. Resilient back propagation algorithm is used to train Multilingual Speech signal. For gender identification and age identification separate models were used.

G.S.Archana et al. [20] proposed a gender classification system using Artificial Neural Network and SVM as classifiers. To train and test these classifiers MFCC, energy entropy and frame energy estimation features were used. These features were extracted from real time male and female voices. They experimentally claimed that SVM classification performed better than ANN in the gender identification by speech for the same feature vector.

Md. Sadek Ali et al. [21] proposed a system for speech encoding, speech analysis, speech synthesis and gender identification. Power spectrum density, frequency at maximum power carry speaker information. First Fourier Transform (FFT) algorithm is used to extract these features. Frequency at maximum power is extracted and threshold value are calculated using which it can be identified gender of a speaker.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

80% accuracy is claimed here. In the literature it is found that till now, no method is obtained for automatic gender identification in robust environment with acceptable accuracy. These are the following reasons for not getting the accurate result of gender identification using speech signal through machine. One of the primary reasons is that every individual's speech characteristics is one of a kind so that makes grouping a hard undertaking. Additionally another test is the commotion calculates. Commotion can be something besides the speaker's voice. These issues are portrayed in more detail beneath:

Every speaker of the dialect is distinctive. The distinction originates from the vocal life structures of speaker. One male and one female's discourse attributes can be fundamentally the same as far as sexual orientation and furthermore individuals from various age gatherings can have comparable discourse attributes as far as age arrangement. So as to get great acknowledgment comes about, the framework must be prepared on bunches of information all together for the framework to be precise.

The biggest hurdle is the noise factor which added with speech in real life environment. The noise can interfere with the actual speech and this can lead to wrong classification. Noise can be anything like, babble noise, street noise, suburban train noise, car noise, restaurant noise or similar kind. So in order to have a reliable gender recognition system, some pre-processing techniques and robust features need to be applied. In this thesis, the focus is mainly on three speech features: pitch and MFCC. This thesis can be differentiated from other works in the way that here a new algorithm is proposed for extracting appropriate frames of speech which are required for gender identification. To see the performance of trained classifiers ELSDSR dataset is used. The algorithms used in this work were less complicated compared to the other works in the literature.

III. CONCLUSION

In this Paper a brief survey is performed for gender identification through Machine using speech signal. Here, it is observed that till now achieved accuracy is not up to the mark for robust environments. The main reason behind this is the non-stationary behavior of speech signals and its additive property with noise signal. Till now, researchers used many speech features like MFCC, RASTA-PLP, LPC, Pitch etc. and various Machine learning algorithms like neural network, SVM, Decision tree, Random Forest etc. to perform the automatic gender identification. Databases like ELSDSR, TIMIT etc. are used for the speech signals.

REFERENCES

1. M. Li, K. Han, and S. Narayanan, "Automatic speaker age and gender recognition using acoustic and prosodic level information fusion," Computer speech and language, Vol. 27, No. 1, pp. 151-167, Jan. 2013
2. H Meinedo and I Trancoso, “Age and Gender Classification Using Fusion of Acoustic and Prosodic Features”, Proc. INTERSPEECH, pp. 2818-2821, 2010
3. R. Nisimura, A. Lee, H. Saruwatari, and K. Shikano, "Public speech-oriented guidance system with adult and child discrimination capability," Proc. ICASSP2004, vol. 1, pp. 433-436, 2004.
4. H. Kim, K. Bae, H. Yoon, “Age and gender classification for a home-robot service” Proc. 16th IEEE International Symposium on Robot and Human Interactive Communication, pp. 122–126, 2007.
5. W. Li, D. J. Kim, C. H. Kim, and K. S. Hong, "Voice-Based Recognition System for Non-Semantics Information by Language and Gender" Electronic Commerce and Security (ISECS), 2010.
6. P. Nguyen, D. Tran, X. Huang, and D. Sharma, "Automatic classification of speaker characteristics" Communications and Electronics (ICCE), 2010.
7. G. Dobry, R. M. Hecht, M. Avigal, and Y. Zigel, "Supervector dimension reduction for efficient speaker age estimation based on the acoustic speech signal." Audio, Speech, and Language Processing, 2011
8. M. H. Bahari, and H. V. Hamme, "Speaker age estimation and gender detection based on supervised non-negative matrix factorization" Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2011.
9. M. H. Sedaaghi, "A comparative study of gender and age classification in speech signals" Iranian Journal of Electrical & Electronic Engineering, 2009.
10. T. Bocklet, G. Stemmer, V. Zeissler, and E. Nöth, "Age and gender recognition based on multiple systems-early vs. late fusion" INTERSPEECH, 2010.
11. M. Feld, F. Burkhardt, and C. A. Müller, "Automatic speaker age and gender recognition in the car for tailoring dialog and mobile services" INTERSPEECH, 2010.
12. F. Metze, J. Ajmera, R. Englert, and U. Bub, "Comparison of four approaches to age and gender recognition for telephone applications" Acoustics, Speech and Signal Processing, 2007.
13. C. A. Müller, F. Wittig, and J. Baus, "Exploiting speech for recognizing elderly users to respond to their special needs" INTERSPEECH, 2003.
14. M. W. Lee, and K. C. Kwak. "Performance Comparison of Gender and Age Group Recognition for Human-Robot Interaction" International Journal of Advanced Computer Science & Applications, 2012.
15. F. Wittig, and C. Müller, "Implicit Feedback for User-Adaptive Systems by Analyzing the Users' Speech.", 2003.
16. T. Bocklet, A. Maier, and E. Nöth, "Age determination of children in preschool and primary school age with GMM-based supervectors and support vector machines/regression" Text, Speech and Dialogue, 2008.
17. M. H. Bahari, and H. V. Hamme, "Speaker age estimation using Hidden Markov Model weight supervectors" Information Science, Signal Processing and their Applications (ISSPA), 2012.
18. S. Khanum and M. Sora, “Speech based Gender Identification using Feed Forward Neural Networks”, International Journal of Computer Applications (0975 – 8887), 201.
19. S. Sharma, A. Shukla and P. Mishra, “Speaker and Gender Identification on Indian Languages using Multilingual Speech ”, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014.
20. G.S. Archana, M. Malleswari and N. Islam, “Gender Identification and Performance Analysis of Speech Signals”, Proceedings of 2015 Global Conference on Communication Technologies (GCCT 2015).
21. Md. S. Ali1, Md. S. Islam and Md. A. Hossain, “GENDER RECOGNITION SYSTEM USING SPEECH SIGNAL”, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.1, February 2012.
22. L. Feng, “Speaker Recognition, Informatics and Mathematical Modelling”, Technical University of Denmark, 2004.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)