IDENTIFYING THE GENDER OF A VOICE USING ACOUSTIC PROPERTIES

Sumant Sahney, Theertha Babu, Kiranmayee Narahari, Abhishek Karan

ABSTRACT

Machine learning is the ability of a computer to learn to make decisions without being programmed explicitly. It has evolved to artificial intelligence (AI). Machine learning focuses on making such computer programs that have the ability to change when in contact with new data. Data prediction and learning from data is the main task of such programs and machine learning helps to explore the study and construction of algorithms which help to do the same. We are witnessing rapid advancements in this field and it is a possibility in the near future that voice interaction systems will replace our normal way of interaction through standard keyboards in the near future. Today, some notable examples in voice interaction systems are Microsoft Kinect and Apple SIRI which perform really well, but, like any other technology, there is a wide scope of improvement in every speech system that is available today. They have their own drawbacks and continuous research is being done to increase the performance of such systems. One method to increase the performance of speech systems is using preprocessing like gender recognition. The paper discusses automatic gender identification systems using acoustic properties of speech. The different samples which have been processed using acoustic analysis and then applied to different machine learning algorithms, to learn gender-specific traits, are being discussed and experimented.

Keywords: *Machine learning, gender recognition, artificial intelligence, voice recognition.*

INTRODUCTION

Our daily lives revolve more and more around computers with each passing day. The significance of computers and machines in our daily life is following an incremental curve, thus making the interaction between humans and machines continually more important. The desire of humans to communicate with machines in a natural way has led to the evolution of language processing. The near future may see the replacement of standard keyboards with voice interaction systems for the ease and comfort of humans. The current advancements in this field are supportive of such claims. State of the art technologies like Microsoft Kinect and Apple SIRI are leading the market developments in this area. They have a huge user base due to their effective working but there are drawbacks in every speech system available today and work is being done continuously to enhance the performance of such systems. A lot can be done to increase the performance of speech systems. One of those ways can be using preprocessing like gender recognition. This project focuses on automatic gender identification systems using speech, that is, the system will identify if the speaker is a male or female using the acoustic properties of the voice of the person. Automatic Gender

Identification using speech of a person has a number of applications in the field of natural language processing. We will design a computer program to model acoustic analysis of voices and speech for determining gender. The model will be constructed using recorded samples of male and female voices, speech, and utterances. The samples are processed using acoustic analysis and then applied to a machine learning algorithm to learn gender-specific traits.

LITERATURE SURVEY

In most of the studies [10–16], the acoustic features used for the gender detection depend on the accurate estimation of the fundamental frequency. The most important and challenging task for such a system is the accurate estimation of the fundamental frequency. Inaccurate estimation of the fundamental frequency may lead to a significant reduction in the accuracy of a gender detection system. Moreover, various traditional speech features such as linear predictive coefficients (LPC), linear predictive cepstral coefficients (LPCC), Mel-frequency cepstral coefficients (MFCC), perceptual linear predictive coefficients (PLP), and relative spectral PLP coefficients (RASTA-PLP) are used in [10, 12–14, 17, 18] for gender detection. The author in the study [19] claimed that it is not necessary that the features used for speech recognition will provide good results for gender detection too. Therefore, it became necessary to explore new features for gender detection other than the regular, traditional speech features, and also make sure that those features should not rely on the accurate estimation of the fundamental frequency.

In this project, we propose a new type of method for automatic gender detection. We will use machine learning algorithms to train using a given dataset and provide better results as compared to existing results in the most appropriate model on which researchers are currently working [2,4-9,20]. For example, two acoustic features, pitch and first formant, are extracted by linear predictive analysis to construct a gender detection system in [1, 17]. The first feature relates to voice source and the second to the vocal tract. Multiple researches are going on to determine the one most distinguishing feature of the male and female voices which can be used for identifying them. The area in which gender is recognized by using machine learning is yet to be explored. Thus we propose an easy and efficient way to determine gender by creating an ensemble which helps us in achieving high accuracy even when we have a small data set [3].

PROPOSED SYSTEM

Determining a person’s gender as male or female, based upon a sample of their voice, seems to initially be an easy task. Humans are capable of detecting the difference between a male or female voice with ease via the human ear. The brain has its own mechanism for differentiating between voices. However, we plan on designing a computer program to do this.

In statistics and machine learning, a single learning algorithm gives a good performance, but to enhance the results, it is better to use ensemble methods which use multiple learning algorithms so that the predictive performance is greater than that of a single algorithm. A machine learning ensemble refers to a definite finite set of alternative models, but generally allows for much more flexible structure to exist among those alternatives. The task of supervised learning algorithms search through a hypothesis space to find a suitable hypothesis that will make good predictions for a particular problem. Ensembles combine multiple hypotheses to form a better hypothesis. Using the same base learner, certain methods generate multiple hypotheses. These methods are known as Ensemble. Different technologies such as Simple Threshold Frequency, CART Model of Voice Acoustics, Random Forest, Boosted Tree, SVM, and XGBoost are used in this project.

PROPOSED ARCHITECTURE

The proposed architecture is as shown in the figure. We have two phases in the implementation of the system. The first is the Training Phase and the second is the Recognition Phase. To implement any Supervised Learning Algorithm, we need to provide the System with Training Data (Feature). In our case, the Training data is the collected samples of voices over telephone conversations. The voices have been already identified and their gender-distinguishing characteristics are noted. These details are extracted by the system from all the sources possible. Data Extraction, Clean-up and Transformation is done by the system to enhance the quality of the extracted data. The data is then stored in the Database.

The second phase is the Recognition Phase. In this phase, sample data is taken, that is, the input speech to be tested is taken and fed to the system. Again the data extraction, cleanup, and transformation process takes place to extract meaningful data. The next step is to start pattern recognition. The Algorithm used is responsible for finding similarities and dissimilarities in the data. The data is sorted, searched through, and any meaningful patterns that are possibly available are recognized by the system. The design logic then helps to build the decision tree or other classification patterns which is defined by the Decision Logic. The given logic of the algorithm helps to determine, stepwise, the answer to our problem. The final decision tells us whether the voice sample is that of a male or a female.

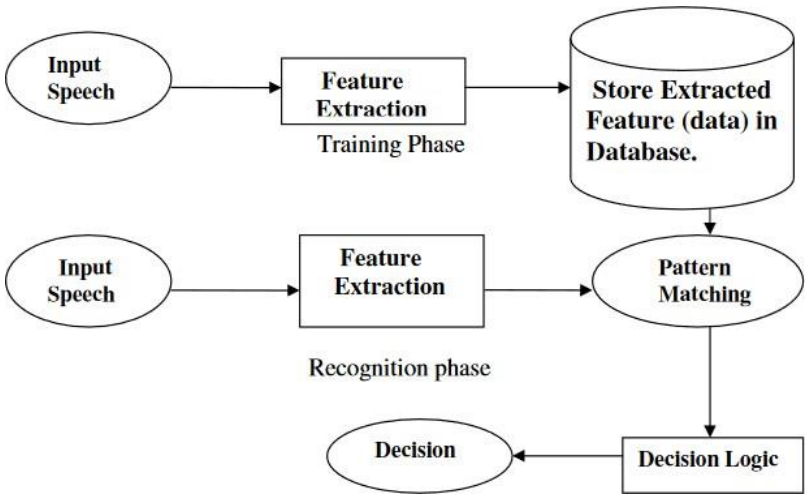

Fig 1. Architectural Diagram Implementation And Expected Result

The first step to implement the model is to divide the speech in half to allow automatic speaker recognition, which reduces computations and enhances the speed of the system. Next, speaker adaptation needs to be enhanced as part of an automatic speech recognition system. At the same time, it is imperative to sort telephone calls by gender for gender sensitive surveys. This will help us to identify the gender and remove the gender specific components.

By achieving higher compression rates of a speech signal, we can enhance the information content to be transmitted and also save the bandwidth. We plan on stacking together multiple models with each model's outputs as a classification of male or female, based upon the audio file. We will test all the models first for their efficiency and then choose the best among the models such as CART, Random Forest, Boosted Tree, SVM etc. We will take the output of the models and then feed them into another model and figure out which weighs higher and hopefully increase the accuracy. We plan to achieve an accuracy of 95%+.

The following algorithms were implemented by us for the project:

Baseline Algorithm: It is a simple algorithm which tends to interpret all samples as ‘male’ voices irrespective of the acoustic properties of the voice.

Logistic Regression: We are using R to create a regression model with 15 statistical properties, which helps to increase efficiency.

Classification and Regression Tree: CART tree uses the mode frequency (mode) to determine the root node for identifying the gender. Next, it checks the minimum fundamental frequency, along with more specific properties, such as maximum dominant frequency, first quantile hertz, skewness, median frequency, and detailed mode frequency.

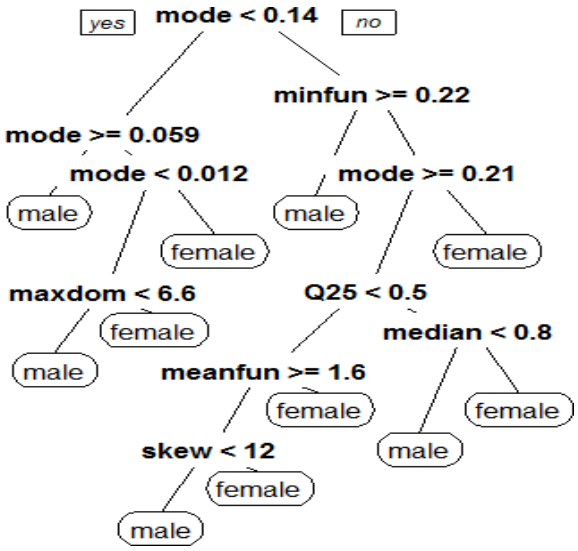

Fig 2: CART model

Random Forest: It is similar to the CART tree. The difference is that in the Random Forest Algorithm, multiple CART trees are built by bootstrapping the CART algorithm.

XG Boost: XG Boost uses a gradient booster and is a combination of linear model predictor and decision tree maker. It uses both in parallel computation which makes it faster and more efficient than the normal algorithms.

Stacked Ensemble: Stacked Ensemble Algorithm is an ensemble or combination of more than one algorithm to increase accuracy and efficiency. It boosts accuracy more than any one individual algorithm. In our project, we have combined the SVM, Random Forest and XG Boost Algorithms. The outputs from the individual algorithms are taken and fed into yet another classification tree to increase accuracy.

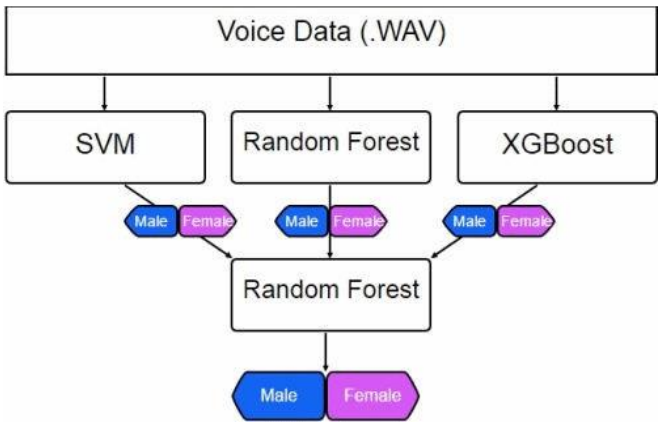

Fig 3: Stacked ensemble

CONCLUSIONS

We took 3,168 voice samples of various males and females to train the model. When we tested with frequencies outside the range 0-280 Hz, Stack Ensemble Algorithm gave the best results, but when we tested with frequencies in the above range, XG Boost Algorithm proved to be the best with an efficiency of 98%.

The accuracies of the various Algorithms are as follows (the percentages denote the accuracy on the Training Set and the accuracy on the Test set respectively):

Baseline Algorithm 50%/50%
Logistic Regression 61%/59%
Classification and Regression Tree (CART) 72%/71%
Random Forest 81%/78%
Stacked Ensemble 89%/89%
XG Boost 91%/84%
Generalized Boosted Tree Regression 100%/87%
XGBoost (Updated with frequency range 0-280 Hz) 100%/99%

REFERENCES

[1] H.Harb and L.Chen, "Voice-based gender identification in multimedia applications," Journal of Intelligent Information Systems, vol. 24, no. 2, pp. 179–198, 2005. View at Publisher View at Google Scholar · View at Scopus

[2] M. Shamim Hossain and G. Muhammad, "Cloud-assisted Industrial Internet of Things (IIoT)-enabled framework for health monitoring," Computer Networks, 2016. View at Publisher · View at Google Scholar

[3]G.Muhammad, "Automatic speech recognition using interlaced derivative pattern for cloud based healthcare system," Cluster Computing, vol. 18, no. 2, pp. 795–802, 2015. View at Publisher · View at Google Scholar · View at Scopus

[4] M. Shamim Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib, "Audio- visual emotion recognition using big data towards 5G," Mobile Networks and Applications, 2016. View at Publisher · View at Google Scholar

[5] M. S. Hossain, "Cloud-supported cyber- physical localization framework for patients monitoring," IEEE Systems Journal, 2015. View at Publisher · View at Google Scholar

[6] G. Muhammad, T. A. Mesallam, K. H. Malki, M. Farahat, M. Alsulaiman, and M. Bukhari, "Formant analysis in dysphonic patients and automatic Arabic digit speech recognition," BioMedical Engineering Online, vol. 10, article 41, 2011. View at Publisher · View at Google Scholar · View at Scopus

[7] G. Muhammad, M. AlSulaiman, A. Mahmood, and Z. Ali, "Automatic voice disorder classification using vowel formants," in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '11), pp. 1–6, Barcelona, Spain, July 2011.

[8] M. Bouchayer, G. Cornut, E. Witzig, R. Loire, J. B. Roch, and R. W. Bastian, "Epidermoid cysts, sulci, and mucosal bridges of the true vocal cord: a report of 157 cases," The Laryngoscope, vol. 95, no. 9, pp. 1087–1094, 1985. View at Google Scholar · View at Scopus

[9] M. M. Johns, "Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts," Current Opinion in Otolaryngology and Head and Neck Surgery, vol. 11, no. 6, pp. 456–461, 2003. View at Publisher · View at Google Scholar · View at Scopus

[10] K. Wu and D. G. Childers, "Gender recognition from speech. Part I: coarse analysis," Journal of the Acoustical Society of America, vol. 90, no. 4 I, pp. 1828–1840, 1991. View at Publisher · View at Google Scholar · View at Scopus

[11] S. M. R. Azghadi, M. R. Bonyadi, and H.Sliahhosseini, "Gender classification based on feedforward backpropagation neural network," in Artificial Intelligence and Innovations 2007: From Theory to Applications: Proceedings of the 4th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2007), C. Boukis, L. Pnevmatikakis, and L. Polymenakos, Eds., vol.
247 of IFIP The International Federation for Information Processing, pp. 299–304, Springer, Berlin,

Germany, 2007.View at Publisher · View at Google Scholar

[12] S. Gaikwad, B. Gawali, and S. C. Mehrotra, "Gender identification using SVM with combination of MFCC," Advances in Computational Research, vol. 4, no. 1, pp. 69– 73, 2012. View at Google Scholar

[13] M. Pronobis and M. Magimai-Doss, "Analysis of F0 and cepstral features for robust automatic gender recognition," Tech. Rep. Idiap-RR-30-2009, Idiap, 2009. View at Google Scholar

[14] Y.-M. Zeng, Z.-Y. Wu, T. Falk, and W.-Y. Chan, "Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech," in Proceedings of the International Conference on Machine Learning and Cybernetics, pp. 3376–3379, Dalian, China, August 2006. View at Publisher · View at Google Scholar · View at Scopus

[15] G. Chen, X. Feng, Y. Shue, and A. Alwan, "On using voice source measures in automatic gender classification of children's speech," in Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH '10), pp. 673–676, Chiba, Japan, 2010.

[16] F. Lingenfelser, J. Wagner, T. Vogt, J. Kim, and E. André, "Age and gender classification from speech using decision level fusion and ensemble based techniques," in Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH '10), pp. 2798–2801, Chiba, Japan, September 2010.

[17] K. Rakesh, S. Dutta, and K. Shama, "Gender recognition using speech processing techniques in labview," International Journal of Advances in Engineering & Technology, vol. 1, no. 2, pp. 51–63, 2011. View at Google Scholar

[18] M. Sigmund, "Gender distinction using short segments of speech signal," International Journal of Computer Science and Network Security, vol. 8, no. 10, pp. 159–162, 2008. View at Google Scholar

[19] D. S. Deiv, Gaurav, and M. Bhattacharya, "Automatic gender identification for hindi speech recognition," International Journal of Computer Applications, vol. 31, no. 5, pp. 1–8, 2011. View at Google Scholar

[20] V. N. Sorokin and I. S. Makarov, "Gender recognition from vocal source," Acoustical Physics, vol. 54, no. 4, pp. 571–578, 2008. View at Publisher · View at Google Scholar · View at Scopus