An Analysis of Automatic Voice Recognition and Speaker Identification Algorithms and its Applications

Ram Sethuraman[1], J.Selvin Paul Peter[2], Shanthan Reddy Middela[3]

[1,3]*Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, India*
[2]*Faculty of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, India*
*Corresponding Author Email: Ramsethuraman123@gmail.com

Abstract

Voice recognition is the domain which is used to identify the speaker behind a speech through their voice. In the field of research, Voice recognition is a domain which has been widely explored by data mining experts and used for various applications. The features of the voice are extracted through methods like MFCC and then various Data Mining and Machine learning algorithms are applied for each specific application. Researchers have explored and tested the efficiencies of various algorithms for various purposes. There appears to be specific algorithms which outperform the rest in certain applications whereas they tend to perform badly for certain other applications. This paper aims to discuss the various Voice recognition techniques and its uses in various domains. The work aims in providing the characteristics and limitations of these approaches.

Keywords: MFCC; Voice Recognition; Speaker Recognition; Speaker Identification; Applications.

1. Introduction

Speech is the most widely used method of Human Communication. There is a lot of confusion on the difference between speech recognition and Voice recognition. Both, speech recognition and Voice recognition use recordings of human voice. Speech recognition does not care about the identity or voice of the person. It is only used to detect the words and the meaning behind these words. Whereas, Voice recognition is the exact opposite of this. It is used to detect the speaker through the voice and it disregards the language or the meaning behind the language. Voice recognition is used in various domains. Like fingerprints, voices are used as a unique identity called voice print. The most eminent usage of Voice recognition is in Biometric systems. A voiceprint is used as a secure method of validating one's identity unlike passwords which can be stolen. But this is not the only usage of Voice recognition algorithms. The methods used for this purpose has evolved over time. It started with rule-based approaches which were used a few decades ago, which was superseded by traditional machine learning algorithms which clearly outperformed the older methods. Now the trend is shifting towards Deep learning and neural network models for this purpose. This transition has been slow, but there has been a clear breakthrough in this field after the popularity of machine learning and deep learning.

2. Related Works

Over the past few years. Voice recognition algorithms are being used in a wide variety of fields like Emotional analysis, Disease detection, Character identification without Script, Detection of age from voice, Detection of gender from their voice, detection of the type of animal and their breed through their noises, Detecting the error in electrical appliances through their noise and many more.

3. Voice Recognition Algorithms for Various Domains

The voice recognition algorithms have been evolving over time in various domains. The most notable work in the fields have been studied and presented. We have classified six major fields where the algorithms are used and the most efficient out of all of them.

A. Detection of Human Emotions

Human emotions can be easily understood through their voice. This can be used in various fields to add another layer of intelligence to the existing systems. Having a machine to understand your emotions is extremely powerful. We have been able to achieve more than 98% accuracy with advancements in Voice recognition algorithms.

T.Seehapoch[1] has computed has computed speech emotions from three different databases. MFCC is used to extract the features and support vector machines(SVM) is used as a classification model. It has consistently given around 90-93% accuracy. But this has given a higher accuracy of 98% in Thai emotions database.

$$W(\alpha) = \sum_{i=1}^{N} \alpha_i - 1/2 \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_k)..[1]$$

E.H.Kim[2] designed an algorithm to recognize emotions even in a noisy environment. He used Eigen-Fast Fourier Transform(FFT) method for this purpose. The accuracy was about 90%

The disadvantages of the methods which were proposed above are that the features were analyzed individually and then combined into a feature space. The combination of each feature representing the emotions are not considered.

To overcome these difficulties Sathit Prasomphan[3] proposed an approach using speech spectrogram. The increasing popularity of

deep learning was utilized well here. The classifier used was Neural networks. The accuracy was consistently above 98%

B. Disease detection

This is currently a domain where a lot of research has been going on. Researchers have been able to find a relation between diseases and voices of a person. This approach can be used to detect diseases at an early stage.

Vishakha Pareek[4] proposes an algorithm to detect Coronary heart disease from analysing the voice of the person and how it changes over time. They have been able to do this through analysis of various voice parameters such as jitter and shimmer.

Shim and Jita are period to period variation of amplitude and pitch respectively.

$$Jita = \frac{1}{N-1} \sum_{i=1}^{N-1} |T^i - T^{i+1}|$$

$$Shim = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A^i - A^{i+1}|}{\frac{1}{N} \sum_{i=1}^{N} A^i}$$

T(i), i=1,2…N = Extracted period to period pitch A(i), i=1,2…N = Extracted peak to peak amplitude

C. Automatic Character Identification without Script

Character Identification is a method which is used to identify the character who is talking. This is done especially in movies where speech recognition can detect only the Script. The Character can be identified automatically using Voice recognition algorithms.

Arsha Nagrani[5] has proposed a method to use speaker identification algorithm combined with face recognition to build a highly accurate system to identify the characters. They have taken a TV series called sherlock and have identified the voice of the person automatically without the script. This will prove to be a way to recognize speaker in movies without manual intervention.

D. Age and Gender prediction

Age and Gender recognition using voice features is a relatively new area of research. Jiri Pribil[6] has proposed a novel method using Gaussian mixture models(GMM) for classification of their Age and Gender.

For Gender, they have used three classifications. Male, Female and Child. Children under 12 have a different voice and thus they were considered as another classification. This does not predict the exact age but uses four classes. Child, Young, Adult and Senior. Surprisingly this method has shown a high accuracy. Accuracy was 95% for Child, 99.1% for male and 98.1% for female. The average overall accuracy of the GMM based system was 92.3% which is higher than any other approach in this field.

E. Detection of Animals and their Breeds

Detection of Animals using their voice has been successful over the past few years. Scientists have estimated that there are over 30 million animals which are divided into six major groups namely invertebrates, fishes, amphibians, reptiles, mammals and birds. Among these groups, there are around 5000 species of mammals alive today. This makes it an extremely difficult task to identify the voice of animals manually. Researchers lately, are even trying to detect even the specific breeds of each animal with their voices. One example of this would be a system proposed by Che Yong Yeo[7] on Dog Voice identification for Detection System. They have used MFCC and Dynamic Time Warping(DTW) algorithms to detect the animal voices. They have been able to achieve an average accuracy of 80%.

F. Detecting the Error in Electrical Appliances Through Noise

The error in an electric appliance can be detected in an early stage through voice recognition algorithms. The exact details and cause of the fault can be analysed before further damage. Mohamed Fezari[8] has proposed a method for early detecting and classification faults in Rotating machines using Euclidean Distance(ED) based approach and K-Nearest Neighbour(K-NN) based approach. The accuracy of the Euclidean based approach was 92% and that of K-NN was 94%

4. Conclusion

In this survey paper we have presented an extensive survey of voice recognition algorithms. We have presented a study on status of research on voice recognition and speaker identification. Several speaker recognition algorithms have been used in various domains. We have concisely given the best and the most recent applications by comparing the accuracy of various works. This work helps in realizing the wide usage of voice recognition algorithms in various fields and encourages to explore other fields where they can be used.

References

[1]   T. seehapoch and S. Wongthanavasu, "Speech emotion recognition using Support Vector Machines", IEEE Int. conference on knowledge and smart technology (KST), 2013.

[2]   Kim EH, Hyun KH, Kim SH, Kwak YK. Improved emotion recognition with a novel speaker-independent feature. IEEE/ASME Transactions on Mechatronics. 2009; 14(3):317–25.

[3]   S. Prasomphan. Detecting human emotion via speech recognition by using speech spectrogram 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA) pp. 1-10 2015.

[4]   Vishakha Pareek, Sharma RK. Coronary heart disease detection from voice analysis. IEEE students conference on electrical, electronics and computer science; 2016

[5]   Arsha Nagrani, Andrew Zisserman. From Benedict Cumberbatch to Sherlock Holmes: Character Identification in TV series; 2018

[6]   Pribil, J., Pribilova, A., Matousek, J. (2016). GMM-based speaker gender and age classification after voice conversion. In First international workshop on sensing, processing and learning for intelligent machines (SPLINE), Denmark.

[7]   C.Y. Yeo, S.A.R. Al-Haddad, C.K. Ng. Dog voice identification (ID) for detection system. Second International Conference Digital Information Processing and Communications; (2012), pp. 120-123

[8]   M. Fezari, F. Z. Taif, M. M. Lafifi, and B. Boulebtateche, Noise emission analysis a way for early detection and classification faults in rotating machines, in Power Electronics and Motion Control Conference and Exposition (PEMC), 2014 16th International, 2014, pp. 1094-1099.

[9]   S.V.Manikanthan and D.Sugandhi " Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel " International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume- 7 ,Issue 1 –MARCH 2014.

[10] T. Padmapriya and V. Saminadan, "Inter-cell Load Balancing technique for multi-class traffic in MIMO-LTE-A Networks", International Journal of Electrical, Electronics and Data Communication (IJEEDC), ISSN: 2320- 2084, vol.3, no.8, pp. 22-26, Aug 2015.

[11] S.V.Manikanthan and K.Baskaran "Low Cost VLSI Design Implementation of Sorting Network for ACSFD in Wireless Sensor Network", CiiT International Journal of Programmable Device Circuits and Systems,Print: ISSN 0974 – 973X & Online: ISSN 0974 – 9624, Issue : November 2011, PDCS112011008.