SPEAKER DEPENDENT SPEAKER RECOGNITION USING SVM AND HMM

Y. PRASANNA KUMAR

M.TECH STUDENT DEPT. OF ECE, ANU [1]

CH. D. UMASANKAR

ASST. PROFESSOR DEPT. OF ECE, ANU[2]

[1]Acharya Nagarjuna University college of engineering and technology, Guntur, A.P, (522510), INDIA

[2]Acharya Nagarjuna University college of engineering and technology, Guntur, A.P, (522510), INDIA

prasannakumaryalla07@gamil.com[1]

Abstract

Speaker recognition is the process of recognizing the speaker based on characteristics such as pitch, tone in the speech wave.Background noise influences the overall efficiency of speaker recognition system and is still considered as one of the most challenging issue in Speaker Recognition System (SRS). Support Vector Machine (SVM) and Hidden Markov Model (HMM) are widely used techniques for speech recognition system. Acoustic features like Mel Frequency Cepstral Coefficients (MFCC) are extracted. Finally a speaker recognition using MFCC for feature extraction and SVM and HMM for classification then compares the two techniques on the basis of identification rate.

Keywords: MFCC, SVM, HMM, Speaker recognition, Classification

1. INTRODUCTION

Speaker identification has been the subject of active research for many years, and has many potential applications where propriety of information is a concern. Speaker identification is the process of automatically recognizing a speaker by machine using the speaker's voice. The most common application of speaker identification systems is in access control, for example, access to a room or privileged information over the telephone. Also it has a very useful usage for speaker adaptation in automatic speech recognition system. Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance.

Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker. Speaker recognition methods can also be divided into speaker-independent and speaker-dependent methods. In a speaker-independent system, speaker models capture characteristics of somebody's speech, which show up irrespective of what one is saying. In a speaker-dependent system, on the other hand, the recognition of the speaker's identity is based on his or her speaking one or more specific phrases, like passwords, card numbers, PIN codes, etc. The choice of which technology to use is application-specific

The objective of this work is to design an efficient system for human speech recognition that is able to identify and verify human speech more accurately. This work presents a technique of speaker-dependent speaker identification using MFCC and Support Vector Machine (SVM) and Hidden Markov Model. Mel-frequency cepstrum coefficients (MFCCs) and their statistical distribution properties are used as features, which will be inputs to the classifiers.

2. SPEAKER RECOGNITION

Speaker recognition has traditionally relied on low-level acoustic features from speech signals. This has typically been done using a text-independent, bag-of-frames approach, where each signal frame is treated separately and independent of other frames as well as speech text. Gaussian mixture models are typically built from acoustic features, and used for classification of speakers.

New approaches to speaker and background model training have given rise to many recent developments in speaker recognition. Recently, various text-dependent approaches have surfaced, including a keyword Hidden Markov Models (HMM) approach. This approach also deviates from the traditional bag-offrames approach by taking into account relationships in time among acoustic features for different signal frames. In many of these text-dependent approaches, acoustic features are obtained not for the entire speech signal, but for only parts of the signal corresponding to certain words. The words are usually chosen rather arbitrarily – usually based on perceived frequency and variability in pronunciation. The belief is that the different ways that people pronounce these words will provide enough speaker discriminative information.

This paper discusses two speaker-dependent speaker recognition approaches that were implemented. The first is the traditional keyword HMM approach (which will act as a baseline for comparison), while the second is a another approach using support vector machine (SVM) models trained using data from HMMs

3. FEATURE EXTRACTION

3.1Mel Frequency Cepstral Coefficients

MFCCs are short-term spectral features and are widely used in the area of audio and speech processing. The MFCCs have been applied in a range of speech recognition and have shown good performance compared to other features. The low orders MFCCs contain information of the slowly changing spectral envelope while the higher orders MFCCs explain the fast variations of the envelope. MFCC’s are based on the known variation of the human ear’s critical bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies to capture the phonetically important characteristics of speech and audio. To obtain MFCCs, the speech signals are segmented and windowed into short frames of 20 msec.

4. CLASSIFIERS

(a)Support Vector Machine

Support vector machine (SVM) is a kernel-based technique which is successfully applied in the pattern recognition area and, is based on the principle of structural risk minimization (SRM) [8]. SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors [9]. If the data are linearly separated, SVM trains linear machines for an optimal hyperplane that separates the data without error and into the maximum distance between the hyperplane and the closest training points. The training points that are closest to the optimal separating hyperplane are called support vectors. Through some nonlinear mapping SVM maps the input patterns into a higher dimensional feature space. SVM generally applies to linear boundaries. Three types of SVM kernel functions were used in this work are Polynomial, Gaussian and Sigmoidal respectively.

(b)Hidden Markov Model

HMM is powerful statistical tool which is widely used in pattern recognition. Especially, the HMM has been developed extensively in speech recognition system over the last three decades. There are two main reasons for choosing HMM in speech processing. First, the transition and duration parameters in HMM may properly reveal the evolution of features over time, which is very important in modeling speech audio perception. Second, there are many kinds of variations of the HMM as well as experiences of using them which are developed in speech recognition researches. This makes HMM a mature technique to be applied in this research.

4. PROCEDURE

Training Procedure:

1. Transition probability and emission probability is randomly initialized based on the observation (0=1)

2. Based on the randomized emission and transition probability the model is trained for the given word feature vector. SVM and HMM Modeling Techniques for Speech Recognition Using and MFCC

3. The new transition and emission probability are estimated based on the word vector.

4. The HMM for the individual word is represented by the parameters namely mean, co-variance, mixmat, emission, transition and LL.

Testing Procedure:

1. Loglikely hood ratio for each and every frame in the isolated word is computed against every model. The most likely model is considered as a winner.

2. Based on the frequency of winning HMM models the most occurring model is considered as the Markov model representing that word's feat vector.

Comparison

Finally a speaker recognition using MFCC for feature extraction and SVM and HMM for classification then compares the two techniques on the basis of identification rate.

5. RESULTS

Technique	Identification rate
MFCC & SVM Without noise	66%
MFCC & SVM With noise	50%
MFCC & HMM Without noise	83%
MFCC & HMM With noise	75%

Table 1: Identification rate with and without noise for MFCC & SVM

Technique	Speaker - A (Chaithu)			Speaker - B (Sampath)			Speaker -C (Naveen)			Speaker - D (Abbu)		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
MFCC & SVM without noise	1	_	1	1	1	1	_	_	_	1	1	1
MFCC & SVM with noise	1	_	1	_	1	_	1	_	_	1	_	1
MFCC &HMM Without noise	1	1	1	_	1	_	1	1	1	1	1	1
MFCC &HMM With noise	1	1	1	_	1	_	1	_	1	1	1	1

Table 2: Identification rate with and without noise for MFCC & SVM (Different speakers)

Here in the above table '1' indicates correct recognition and '_' indicates wrong recognition.

6. CONCLUSION

A system has been developed for speaker recognition using SVM and HMM modeling techniques MFCC features are extracted. Features for each speech sample are extracted and those models were trained successfully. SVM and HMM were used to model each individual utterance. From the exhaustive analysis, SVM shows an accuracy of 66% for MFCC and HMM shows an accuracy of 83% using MFCC. Finally a speaker recognition using MFCC for feature extraction and SVM and HMM for classification then compares the two techniques on the basis of identification rate.

REFERENCES

1. Anusuya, M.A., Katti, S.K.: Front end analysis of speech recognition: A review. Int. J. Speech Technology 14, 99–145 (2011)

2. Ramírez, J., Górriz, J.M., Segura, J.C.: Voice Activity Detection- Fundamentals and Speech Recognition System Robustness. Robust Speech Recognition and Understanding, 1–22 (2007) ISBN 978-3-902613-08-0

3. Das, B.P., Parekh, R.: Recognition of Isolated words using features based on LPC, MFCC, ZCR and STE with Neural Network Classifiers. International Journal of Modern Engineering Research (IJMER) 2(3), 854–858 (2012)

4. Herscher, M.B., Cox, R.B.: An adaptive isolated word recognition system. In: Proc. Conf. on Speech

Communication and Processing, Newton, MA, pp. 89–92 (1972)

5. Thiang, W.S.: Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile root. In: International Conference on Information and Electronics Engineering (IPCSIT), vol. 6, pp. 179–183 (2011)

6. Bhattacharjee, U.: A comparative Study of LPCC and MFCC Features for the Recognition of Assamese Phonemes. International Journal of Engineering Research and Technology (IJERT) 2(1), 1–6 (2013)

7. Zolnay, A., Schulueter, R., Ney, H.: Acoustic feature combination for robust speech recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 457–460 (2005)

8. Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S.: Choosing multiple parameters for support vector machines. Machine Learning 46, 131–159 (2002)

9. Duan, K., Keerthi, S., Poo, A.: Evaluation of simple performance measures for tuning SVM hyperparameters. Neuro Computing 51, 41–59 (2003)

10. Bourouba, H., Bedda, M., Djemil, R.: Isolated words recognition system based on Hybrid approach DTW/GHMM. Informatica 30, 373–384 (2006)

11. Baum, L.E.: An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Equalities 3, 1–8 (1972)

12. Reynals, S., Morgan, N., Bourland, H., Franco, R.: Connectionist probability estimators in HMM Speech Recognition. IEEE Trans. on Speech and Audio Processing 2(1), 161–174 (1994)