IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Appraisal of Hidden Markov Model and Dynamic Time Wrap Techniques for Isolated Word in Speech Recognition

Indu Bala[1], Dr. Kailash Bahl[2]

Pursuing M.Tech Degree in Computer Science and Engineering at PIET, Patiala, Punjab, India[1]

Professor Computer Science and Engineering at PIET Computer Science and Engineering at PIET,

Patiala, Punjab, India[2]

Abstract: This paper defines the overall review of speech recognition Technique. Speech recognition has the ability of a machine to identify the words in spoken language or convert it into machine readable form. In a real intelligent computer qualities should have that the machine can hear, understand, and act upon spoken information, and also speak to complete the information exchange The objective of this review paper is to sum up some well-known methods used in various stages of speech recognition system such as Dynamic Time Wrap Model and Hidden Markov Model approaches for isolated speech recognition.

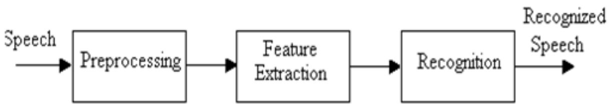
Keywords: Dynamic Time Wrap Model, Hidden Markov Model, isolated speech recognition, time-dependent,, autocorrelation, cepstral, probabilistic model

INTRODUCTION

Speech is the simplest and effective way to exchange information for human beings. While communicating with machines various speech recognition applications based upon different techniques are used. The problem is that the speech recognition software has a limited vocabulary. They can identify only those words which are available in its word list and the voice which is very clear. Speech recognition process allowed human to speak with computer, having a computer to recognize what user is saying and doing this is in real time.

Speech recognition has gained a lot of interest in the researchers from various fields. From various fields, speech recognition has gained a lot of interest in the researchers. Despite(contempt) this, speech recognition has been one of the most difficult problems to solve. Speech recognition is nowadays regarded by market projections as one of the more promising technologies of the future. Voiced command applications are expected to cover many of the aspects of our future daily life. Car computers, telephones and general appliances may be the more likely candidates for this revolution. It may drastically reduce the use of keyboards. . Therefore, speech recognition is important for a computer to reach the goal of natural human-computer communication.

STRUCTURE OF BASIC SPEECH RECOGNITION SYSTEM:



SPEECH CAPTURING

With the help of microphone we can capture the input speech. The analog signal changes into Digital signal in computer with the help of sound card.

The sound card can record the sound and can also play it. Sample size and the sampling frequency can be control while Using Window's MCI (Media Control Interface) commands.

PREPROCESSING

The speech is available in continuous samples After sound-capturing process is complete. Now our next step is pre-process these samples to make available for portent expulsion and esteem. It involves following steps.

1 Background Noise and Silence Removing
2 Preemphasis filter
3 Blocking into Frames
4 Windowing

FEATURE EXTRACTION

Feature extraction is the stages of parameterization of the speech i.e evaluation of speech statement in term of sign helm (feature vectors) which can be used for the recognition purpose. These sign helms should not change with the speaker i.e. the signs should be same for the same statements by different speakers. By using separate methods these features can be summarized for example digital filter, Fourier Transformation or Linear Predictive Coding. Linear Predictive Coding is the most powerful speech feature extraction (e.g., autocorrelation, cepstral coefficient etc.) technique.

DOI 10.17148/IJARCCE.2016.51123

RECOGNITION

This phase is divided into two parts:

- Training
- Testing

As the learning process of a baby goes on same as The training aspect of a recognition system works. A child should experience incidence many times and with a wide mutability before being able to recognize it. The current speech recognition technology does not allow real-time implementation of models comparable to human complexity. This means that the variability of speech must be limited to achieve proper results. To achieve reasonable results the variability of speech must be limited. An unknown statement is scored over indicated patterns in the testing phase. The word recognized is the word corresponding to the reference pattern closest to the unknown pattern.

MODELS OF SPEECH RECOGNITION SYSTEM

1. DYNAMIC TIME WRAP (DTW)

DTW is a method that allows a computer to find an optimal match between two given (time-dependent) sequences under certain restriction. In order to understand DTW, two concepts need to be dealt with:

- Features- The information in each signal has to be represented in which manner.
- Distances- In order to obtain a match path which form of metric has to be used.
- There are two types of distances:
- Local Distance: The computational difference between a feature of one signal and a feature of the other is called Local distance.
- Global Distance: The overall computational difference between an entire signal and another signal of possibly different length is called Global distance.

A) SYMMETRICAL DTW

Speech is a time-dependent process. To find a global distance between two speeches a time alignment must be performed. The best matching pattern is the one for which there is the lowest distance path aligning the input pattern to the former. All possible paths are being evaluated- but this is extremely inefficient as the number of possible paths is exponential in length of the input instead consideration what constraints can impose on the matching process.

Matching paths cannot go backwards in time;
Every frame in the input must be used in a matching path;
Combine local distance scores by adding to give global distance.

B) ASYMMETRICAL DTW:

Only and only one input pattern is used in each frame. This means that it deals with former-length normalization and it is not required to add the local distance in twice for diagonal path transitions. This approach is referred to as asymmetric dynamic programming

HIDDEN MARKOV MODEL

It is a mathematical viewpoint to recognize speech. It is a doubly embedded stochastic process with an underlying stochastic process that is not directly observable (it is hidden) but can be observed only through another set of stochastic processes that produce the sequence of observations. Stochastic modeling entails the use of probabilistic models to deal with uncertain or incomplete information. Confusable sounds, speaker variability s, contextual effects, and homophones words are sources to arise uncertainty and incompleteness in speech recognition.

Thus, stochastic models are particularly suitable approach to speech recognition. Hidden Markov Model is a collection of states connected by transitions. Each transition carries two sets of probabilities: Transition Probability: Which provides the probability for taking this transition, Output Probability: Which defines the conditional probability of emitting each output symbol from a finite alphabet given that a transition is taken. Problems of HMM Evaluation, Decoding Problem, The Learning Problem.

LITERATURE REVIEW OF SPEECH RECOGNITION

The review process was adopted by surveying the research in last few years for extraction of information about some issues. Various research articles were reviewed to cover the review of speech recognition technique. F. Itakura [February 1975] In this found that a new measure of distance for all pole model of speech has been derived on the basis of the likelihood ratio criteria and is applied to automatic recognition of isolated words. An algorithm to find to the best match between the input pattern and a reference pattern is derived. In which the dynamic programming technique is used in conjunction with a sequential decision scheme. The system is implemented on a DDP-516 computer to recognize 200 isolated words. The validity of the scheme has been confirmed experimentally. Further work is in progress to test the system for a greater number of talkers and for telephone connection switched over greater distances.

J. K. Baker [February 1975] This paper describe that termination that the hidden articulatory Markov model as an alternative or companion to standard phone-based HMM models for speech recognition. Found that either in noisy conditions, or when used in tandem with a traditional HMM, a hidden articulatory model can yield improved WER results.

Also shown that the HMM is able to reasonably estimate articulator motion from speech. There are a number of avenues to improve this work. In the future, the plan to add more articulatory knowledge, with rules for phoneme modification that arise as a result of physical limitations and shortcuts in speech production, as was done in (Erler 1996) (for example, vowel nasalization).

DOI 10.17148/IJARCCE.2016.51123

Such rules may help speech recognition systems in the presence of strong co articulation, such as in conversational speech. D. Raj Reddy [April 1976] In this paper stated that the focus has been to review research progress, to indicate the areas of difficulty, why they are difficult, and how they are being solved. The past few years have seen several conceptual and scientific advances in the field. For the first time use the available extensive analysis of connected speech. Know connected speech recognition is not impossible. The role and use of knowledge are better understood.

Almost all systems use knowledge to generate hypotheses and/or verify them. Error and ambiguity can be handled within the framework of search. Stochastic representations and dynamic programming provide a simple and effective solution to the matching problem. F. Jelinek [April 1976] This paper presented that a new approach for visual speech recognition based on a data driven lip model and HMMs. Experiments have demonstrated high recognition performance using very low dimensional shape information only.

The recognition task described is relatively simple because it only consists of four word classes and only deals with isolated words. Nevertheless, recognition tests were speaker independent and have demonstrated high recognition accuracy and generalization ability of the system. More extensive tests with more speakers and sub word classes are necessary to estimate the discrimination ability of shape features for all phonemes. The results are not as good as with 89.58% correct and which was about equivalent to the performance of untrained humans performing the same task. The ability to locate and track lips accurately opens several other potential applications, as example model based image coding, facial animation, facial expression recognition and audio-visual person identification.

Yoseph Linde, Andres Buzo and Robert M. Gray[January 1980] This paper stated that an efficient and intuitive algorithm is presented for the design of vector quantizers based either on a known probabilistic model or on a long training sequence of data. The basic properties of the algorithm are discussed mid demonstrated by examples. Quite general distortion measures and long block lengths are allowed, as exemplified by the design of parameter vector quantizes of tendiensional vectors arising in Linear Predictive Coded (LE) speech compression with a complicated distortion measure arising LPC analysis that does not depend only on the error vector. The hidden-articulator Markov model (HAMM) and have implemented it using HMMs.

Bing-Hwang Juang, David Y. Wong, and H. Augustine, Jr. Gray[April 1982] found that Analytical as well as experimental comparisons between vector and scalar quantization have been presented in detail. It was shown that vector quantization performs a multidimensional clustering process which effectively eliminates unnecessary model spectra. Detailed comparisons between vector and scalar quantization results show that the spectral distortion fluctuates less from frame to frame in vector quantization as compared to scalar quantization.
Lawrence R. Rabiner, and B.H. Juang[January 1986] to search out that to present the theory of hidden Markov models from the simplest concepts (discrete Markov chains) to the most sophisticated models (variable duration, continuous density models).

The purpose to focus on physical explanation of the basic mathematics; hence to avoided long, drawn out proofs and/or derivations of the key results, and concentrated primarily on trying to interpret the meaning of the math, and how it could be implemented to illustrate some applications of the theory of HMMs to simple problems in speech recognition, and pointed out how the techniques could be (and have been) applied to more advanced speech recognition.

Kai-Fu Lee, Hsio-Wuen Hon, and Raj Reddy[January 1990] concluded that SPHINX-a hidden Markov model based system for large-vocabulary speaker-independent continuous speech recognition. On the one hand, HMM's perform better with detailed models. On the other hand, HMM's need considerable training. This need is accentuated in large-vocabulary speaker-independence, and discrete HMM's. However, given a fixed amount of training, model specificity and model trainability pose two incompatible goals.

More specificity usually reduces trainability, and increased trainability usually results in over generality. Joseph W. Picone[September 1993] termination that several popular signal analysis techniques in a common framework that emphasized the importance of accurate spectral analysis and statistical normalization. When viewed in this common framework, the differences amongst these competing approaches seem small when compared to the enormous challenges, still face in the speech recognition problem. All approaches share some important basic attributes: time-derivative information, perceptually motivated transformations, and parameter normalization. L. R. Bahl, P. F. Brown, P. V. de Souza, R. L. Mercer and M.A. Picheny[October 1993] stated that a new technique for constructing Markov models for the acoustic representation of words is described. Word models are constructed from models of sub-word units called fenones. Fenones represent very short speech events, and are obtained automatically through the use of a vector quantizer. The fenonic base form for a word-i.e., the sequence of fenones used to represent the word-is derived automatically from one or more utterances of that word. Since the word models are all composed from a small inventory of sub-word models, training for large-vocabulary speech recognition systems can be accomplished with a small training script. A method for combining phonetic and fenonic models is presented.

DOI 10.17148/IJARCCE.2016.51123

Sahar E. Bou-Ghazale and John H. L. Hansen[May 1998] found that novel modeling approach for speech parameter variations under stress using HMM’s. The variations in overall pitch contour, voiced duration, and overall spectral contour were modeled for angry, loud, and Lombard effect speech. The models were trained with the variations, referred to as perturbations, in speech parameters from neutral to each stressed condition as opposed to training with actual speech parameters. A pitch perturbation model was developed for each stressed condition using a three-state single-mixture Gaussian HMM.

CONCLUSION

In this research we study about the speech recognition, speech capturing, preprocessing and how the speech recognized, for this we study two techniques for speech recognition i.e. Hidden Markov model and dynamic time wrap technique. We try to find the working of both the techniques for speech recognition and on the basis of our study we find that how both the techniques are works and how they recognize the speech and extract the words. We conclude that Hidden Markov model is less expensive in compare to Dynamic time wrap techniques but the performance of Hidden Markov Model is somewhat poorer than the Dynamic Time Wrap based recognizer appears to be primarily because of the insufficiency of the Hidden Markov Model training data. The accuracy of dynamic time wrap technique is more accurate than the hidden Markov model.

REFERENCES

1. F. Itakura, Minimum prediction residual principle applied to speech recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-23, pp. 67-72, February 1975.
2. J. K. Baker, The DRAGON system - An overview, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-23, February 1975, pp. 24-29.
3. D. Raj Reddy, Speech Recognition by Machine: A Review, Proceedings of the IEEE, Vol. 64, No. 4, April 1976, pp. 501-531.
4. F. Jelinek, Continuous Speech Recognition by Statistical Methods, Proceedings of IEEE, Vol. 64, April 1976, pp. 532-556.
5. YosephLinde, Andres Buzo and Robert M. Gray, An Algorithm for Vector Quantizer Design, IEEE Transaction on Communications, Vol. COM-28, No. 1, January 1980, pp. 84-95.
6. Bing-Hwang Juang, David Y. Wong, and H. Augustine, Jr. Gray, Distortion Performance of Vector Quantization for LPC Voice Coding, IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. ASSP-30, No. 2, April 1982, pp. 294-303.
7. Lawrence R. Rabiner, and B.H. Juang, An Introduction to Hidden Markov Models, IEEE ASSP Magazine, Vol. 3, No. 1, January 1986, pp. 4-16.
8. Kai-Fu Lee, Hsio-Wuen Hon, and Raj Reddy, An Overview of the SPHINX Speech Recognition System, IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. ASSP-38, No. 1, January 1990, pp. 35-45
9. Joseph W. Picone, Signal Modeling Techniques in Speech Recognition, Proceedings of the IEEE, Vol. 81, No. 9, September 1993, pp. 1214-1245.
10. L. R. Bahl, P. F. Brown, P. V. de Souza, R. L. Mercer and M.A. Picheny, A Method for the Construction of Acoustic Markov Models for Words, IEEE Transaction on Speech and Audio Processing, Vol. 1, No. 4, October 1993, pp. 443-452.
11. Sahar E. Bou-Ghazale and John H. L. Hansen, “HMM-Based Stressed Speech Modeling with Application to Improved Syndissertation and Recognition of Isolated Speech Under Stress”, IEEE Transaction on Speech and Audio Processing, Vol. 6, No. 3, May 1998, pp. 201-216.

DOI 10.17148/IJARCCE.2016.51123