Journal of Advances and Scholarly Researches in Allied Education
Vol. 20, Issue No. 4, October-2023, ISSN 2230-7540

A Survey of Feature Extraction Techniques for Automatic Speech Recognition System

Shweta Rathore[1]*, Dr. Md. Vaseem Naiyer[2]

[1] Research Scholar, Madhyanchal Professional University Bhopal(M.P.)

Email- rathore.shweta0902@gmail.com

[2] Research Guidde, Madhyanchal Professional University Bhopal(M.P.)

Email- vaseemnaiyer@gmail.com

Abstract - Voice recognition is one of the key evolutions of biometric identification system which not only have identification capability but this technology can be used to create more sophisticated applications. This paper reviews the methodology for extracting features from speech signal. In this paper we have discussed the Components of ASR and Mel Frequency Cepstral Coefficient MFCC,Distance Measurement with Automatic Speech Recognition System(ASR)

Keywords - Biometrics, ASR,Voice recognition,technique

----------X----------

INTRODUCTION

The word “Speech Recognition”, itself explains the meaning that it is the technology developed for device/machine for empowering the understanding of human / non human voice.”ASR is not only a means to provide recognition features in existing system there is more effect of ASR. Let us consider a person not even capable of speck the words properly, suffering a problem for explain what he needs, now from his/her point of view he needs a device that can speak in place of him. This is what exactly the power of ASR technology is. ASR provides the way of research field in which researcher can make devices/machines for this kind of people suffering with speech impairment. So we can say that Automatic Speech Recognition or ASR, as it's known in short, is the technology that allows human beings to use their voices to speak with a computer interface in a way that, in its most sophisticated variations, resembles normal human conversation. Developers across many industries now use automatic speech recognition (ASR) to increase business productivity, application efficiency, and even digital accessibility. This post discusses ASR, how it works, use cases, advancements, and more.

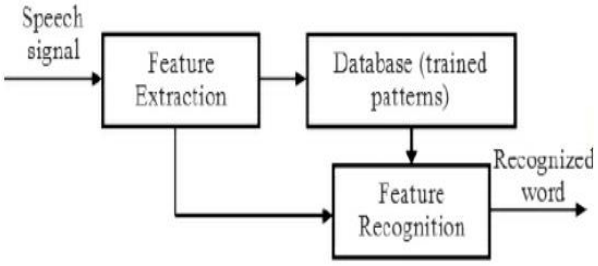

Figure 1: Speech Identification System

COMPONENTS OF ASR

ASR systems are typically composed of three major components —

1- The Lexicon
2- The Acoustic Model
3- The Language Model

that decode an audio signal and provide the most appropriate transcription.

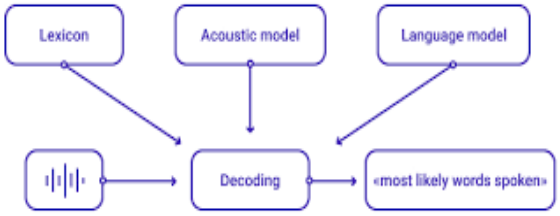

Figure 2: Components of ASR

1. The Lexicon-

A lexicon file enables you to define one or more pronunciations for words that will override the pronunciations provided by the Voice Server. This may be necessary for rare acronyms and foreign words that are incorrect or missing from the base pronunciation dictionary provided with the Voice Server. So a lexicon model specifies the following kinds of information: Types of lexical object and structure of lexical entries. Types of lexical information associated with lexical objects in lexical entries. Relations between lexical objects and structure of the lexicon as a whole lexicon architecture. A key part of any automatic speech recognition system is the lexicon. The lexicon can be tricky to define because it's sometimes used to mean different things depending on the context. In its most basic form, a lexicon is simply a set of words with their pronunciations broken down into phonemes, i.e. units of word pronunciation.

2. The Acoustic Model

One component is an acoustic model, created by taking audio recordings of speech and their transcriptions and then compiling them into statistical representations of the sounds for words. The other component is called a language model, which gives the probabilities of sequences of words. An acoustic model is used in automatic speech recognition to represent the relationship between an audio signal and the phonemes or other linguistic units that make up speech. The model is learned from a set of audio recordings and their corresponding transcripts.

3. The Language Model

Language modeling (LM) is the use of various statistical and probabilistic techniques to determine the probability of a given sequence of words occurring in a sentence. Language models analyze bodies of text data to provide a basis for their word predictions. Language modelling (LM) is the use of various statistical and probabilistic techniques to determine the probability of a given sequence of words occurring in a sentence. Language models analyze bodies of text data to provide a basisfor their word predictions. They are used in natural language processing (NLP) applications, particularly ones that generatetext as an output. Some of theseapplications include , machine translation and question answering.

FEATUREVECTORS AND VECTOR SPACE

If we have set of data representing some especial features for any objective we want to describe, it is very useful to construct a vector of all the data by assigning each data to one component of vector. Let us consider an application, if we think of automatic green house controller system, which will compute the relative humidity and soil moisture of the field. If we measure these two parameters every second, then we can put humidity in first component of vector and soil moisture in second component, we will get a two dimensional vector that is describing the soil moisture and humidity changes in time at fields. As these feature vectors have two components, we can define the vectors in two-dimensional vector space. We can now plot this vector for our calculations. Each point in the graph represents the soil moisture and humidity at any instance of time. Out of these graphs some data is good current conditions for seeds or bad. For mapping this on graph plotted by us we labeled good and bad conditions "+" and "-" Respectively, as shown in Figure 3.

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

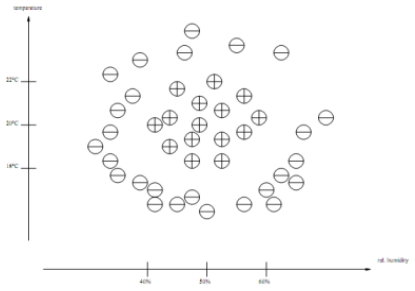

Figure 3: A Feature Vectors Map

Vectors are classified in two types these are as follows.

A. Prototype Vector

This is the simplest way of representing the regions of "comfortable" and "uncomfortable" features. One of the easiest way is to select several of the feature vectors we measured in our experiments for each of our classes (in our example we have only two

classes) and to declare the selected vectors as "prototypes" representing their class.

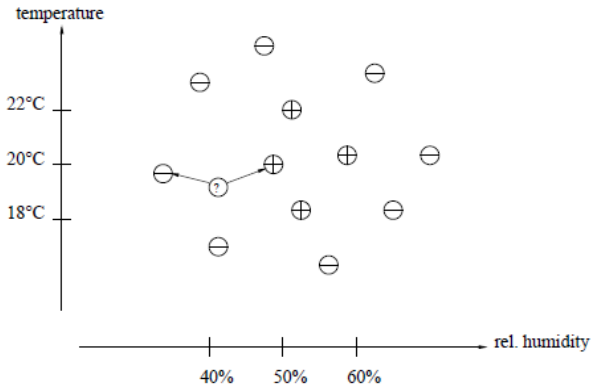

Figure 4: Selected Prototypes

B. Nearest Neighbor Classification

This classification measure the distance between unknown vector to all classes and then assign the unknown vector to the class with the smallest distance. One can also understand it as find the nearest prototype to the unknown vector and assign the unknown vector to the class this "nearest neighbor" represents.

DISTANCE MEASUREMENT

The distance measurement between unknown vector and predefined class. The distance measurement is done by various methodologies here the most important are discussed below.

A. Euclidean Distance

This the standard distance measurement technique to measure the distance between two vectors in feature space. To measure the Euclidean Distance we need to calculate the sum of squares of the differences between the individual components of ~x & p~. mathematical expression as follow.

$$d^2_{Euclid}(\vec{x},\vec{p}) = \sum_{i=0}^{DIM-1}(x_i - p_i)^2 \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$d^2_{Euclid}(\vec{x},\vec{p}) = (\vec{x} - \vec{p})' \cdot (\vec{x} - \vec{p}) \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

This equation can be rewrite as, Where ' denotes the vector transpose, both equation (1) and (2) compute the square of the Euclidean distance, d^2 instead of d. The Euclidean distance is the most commonly used distance measure in pattern recognition.

B. **City Block Distance**

The Euclidean distance involves many multiplication due to the fact that it provides the square of difference of individual terms. So to reduce the calculation steps one can use absolute values of the difference instead of their squares. This is more like a simple street map technique to find the distance between two points. After finding the absolute difference between all points we need to sum up all the values for dimension of the vector space.

C.**Dynamic Time Warping**

The voice/speech signal we have given to system is represented by a number of series of feature vectors computed in 10ms intervals, we are very well known by the fact that number of vector is depends on how fast a user can speak. In speech recognition, we need to classify sequences of vectors. For e.g. if we want to recognize any word or command, for a word "W" whose vectors are "Tx" long, then we will get a vector sequence X̃={x ⃗_0,x ⃗_1,…,x ⃗_(X-1) } from preprocessing stage. What now we need now is to compute distance between known and unknown vector sequences. The computation of distance between "X" and "W" is done by dynamic time warping.

vector sequence to the very class to which the prototype belongs to whose word end grid point was chosen.

Of course, this is just a different (and quite complicated) definition of how we can perform the DTW classification task we already defined in . Therefore, only a verbal description was given and we did not bother with a formal description. However, by the reformulation of the DTW classification we learned a few things:

- The DTW algorithm can be used for real–time computation of the distances
- The classification task has been integrated into the search for the optimal path
- Instead of the accumulated distance, now the optimal path itself is important for the classification task.

CONCLUSION

The most interactive things are those which interconnect modern world application more realistic with real world and the speech recognition and its application is one of the far most selected realistic application of engineering. The speech recognition technology is chosen widely available technology is due to the fact voice communication is the most convenient means of communication between people. This paper attempts to provide a

comprehensive review of speech recognition and methodology for speech recognition.

REFERENCES

4. Zhanyu Ma, Hong Yu, Zheng-Hua Tan And Jun Guo, "Text-Independent Speaker Identification Using the Histogram Transform Model", IEEE Access, VOLUME 4, 2016, pp(9733-9739)
5. K.H.Davis, R.Biddulph, and S.Balashek, Automatic recognition of spoken Digits,,J.Acoust.Soc.Am., 24(6):637-642,1952.
6. H.F.Olson and H.Belar, Phonetic Typewriter , J.Acoust.Soc.Am.,28(6):1072-1081,1956.
7. D.B.Fry, Theoritical Aspects of Mechanical speech Recognition , and P.Denes, The design and Operation of the Mechanical Speech Recognizer at Universtiy College London, J.British Inst. Radio Engr., 19:4,211-299,1959.
8. J.W.Forgie and C.D.Forgie, Results obtained from a vowel recognition computer program , J.A.S.A., 31(11),pp.1480-1489.1959.
9. J.Suzuki and K.Nakata, Recognition of Japanese Vowels Preliminary to the Recognition of Speech , J.Radio Res.Lab37(8):193-212,1961.
10. T.Sakai and S.Doshita, The phonetic typewriter, Information processing 1962 , Proc.IFIP Congress, 1962.
11. K.Nagata, Y.Kato, and S.Chiba, Spoken Digit Recognizer for Japanese Language , NEC Res.Develop., No.6,1963.
12. T.B.Martin, A.L.Nelson, and H.J.Zadell, Speech Recognition b Feature Abstraction Techniques , Tech.Report AL-TDR-64-176,Air Force Avionics Lab,1964.
13. .T.K.Vintsyuk, Speech Discrimination by Dynamic Programming , Kibernetika, 4(2):81-88,Jan.-Feb.1968.
14. H.Sakoe and S.Chiba, Dynamic programming algorithmoptimization for spoken word recognition ,IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-26(1).pp.43-49,1978.

Corresponding Author

Shweta Rathore*

Research Scholar, Madhyanchal Professional University Bhopal(M.P.)