Data Extraction from Web using Speech Recognition

Narendra Kumar Saini[1] Archana Mewara Laxmi[2] Narayan Balai[3]
[1]PG Scholar [2]Assistant Professor [3]H.O.D.
[1,2,3]Department of Electronics & Communication
[1,2,3]Yagyavalkya Institute of Technology, Jaipur, Rajasthan

Abstract— Now days many aspects of speech recognition have been taken over by a deep and smart learning method called Long short-term memory (LSTM). There are many applications of the speech recognition now a days like ticketing system, wheel chair controlling for handicapped, smart data entry, hands free computing, dictation system, smart house appliance control, authentication system, change FM channel while driving a car etc. By use of Speech recognition system any person can handle the system using their voice. The foremost objective of this paper is to extract information from web by making interface between speech and the web. The information is about the data of weather of different cities and stocks of different companies respectively. In this paper we simulate different algorithms of speech processing in MATLAB. For better recognition of speech is performed by the coefficients MFCC (Mel-frequency-Cepstral Coefficients). The efficiency of speech recognition of different words is determined in case of still voice samples. Hidden Markov Model and cross-correlation Techniques are used in this system. The isolated words are record using microphone and simulations of the system done adaptively by using MATLAB. The recognition efficiency is 65% determined by cross-correlation plotting and recognition efficiency 91.60% has been gained by using 400 samples using Hidden Markov method.

Key words: Hidden Markov Model; Mel-Frequency Cepstral Coefficient; Cross-Correlation Method

I. INTRODUCTION

The speech recognition unified the knowledge and research in the field of linguistic in that the technology develop the methodologies in which a person with physical disabilities or who have low vision can communicate to extract the information from web by using their voice. By the use of electronics and computer science technologies the speech digitalization can be useful to develop the communication using spoken language voice signals. The speech signal has multiple characteristics that in change make an impression on intelligibility that comprises enunciation, articulation, rhythm, pitch-rate, fluency etc. The property of speech signal is that for each user the vocal-cord produces different vibrations for different words so there are different frequency bands for each words. For better efficiency the pre-recorded reference signal will be more so that they can match with the speakers voice signal. In Cross-Correlation process, there is need to compare the spectrum of both speaker input signal and reference signals. For feature extraction the Mel-Frequency-Cepstral Coefficients (MFCC) is used for better efficiency instead of Linear Predictive Coding (LPC). The approximation of human auditory system process obtained by using MFC coefficients. The main purposes of these coefficients are to estimate vocal tract system characteristics from the speech signal. The spectrum power obtained onto the Mel-scale. Vector quantitation is a classical quantitation technique which allows the modeling of probability density functions by the distribution of prototype vectors. Hidden Markov models for speech recognition has become predominant for the last several years, as evidenced by the number of published papers and talks at major speech conferences. For better representation of sound before the extraction to develop the quality of speech signals the Hidden-Markov-Model is used for classification of speech signals. Hidden Markov Model (HMM) is a probabilistic pattern matching technique. In probability pattern matching technique the observations are considered to be the output of Stochastic Process and consists of the fundamental Markov-Chain. The Markov Chain Random Process has the property of that the next state depends only on current state of the system. The speech signal is recorded using iPhone-Ear-buds mic 3.5mm. For better training purpose the system in which we have speech signals from 40 persons. The speech signals are recorded in almost calm and noiseless environment for clear recognition. For web data extraction for stock and weather we have used different websites. The objective of this paper is to execute and examine various techniques for speech recognition of the user and which is the best technique for extraction data from web.

II. CROSS CORRELATION ALGORITHM

Cross correlation perform the mathematical relation between reference signal and target signals to know the correlation between them. The mathematical formula for cross correlation is

Rxy= r (m) = $\sum_{n=-\infty}^{\infty} x(n)y(n+m) \quad m = 0, \pm1, \pm2, \pm3$ (1)

Here, x (n) = Reference signal, y (n) = Test signal

III. HIDDEN MARKOV MODEL

Hidden Markov model technology has brought speech recognition system performance to new good levels for a variety of applications so we used Hidden Markov Model is used in speech processing. The main problems handled by Hidden Markov Model are:

1) It helps in determining the probability of the sequence of the observations.
2) HMM helps to foreshow the next observation in the sequence.
3) It helps to find the most likely underlying explanation of the sequence of observation.

Main areas of interest are:

A. *Transition probabilities:*

The conditional distribution $p(z_n \mid z_{n-1})$ is a M x M table A for latent variables with M discrete states, and the marginal distribution p (z_1) describing the initial state is a M vector π

$$p(z_n \mid z_{n-1}, A) = \prod_{m=1}^{M} \prod_{j=1}^{M} A_{jm}^{z_{n-1,j} z_{nm}} \quad (2)$$

$$p(z_1 \mid \pi) = \prod_{m=1}^{M} \pi_m^{z_{1m}} \quad (3)$$

B. Emission probabilities:

From the specific state, the conditional scatterings of the observed variables p $(X_n|Z_n)$

$$p(X_n \mid Z_n, \emptyset) = \prod_{m=1}^{M} p(X_n \mid \emptyset_m)^{z_{nm}} \quad (4)$$

C. Joint Probabilities

$$p(X, Z \mid O) = p(z_n \mid \pi)\left[\prod_{n=1}^{N} p(z_n \mid z_{n-1}, A)\right]\prod_{n-1}^{N} p(X_n \mid Z_n, \emptyset) \quad (5)$$

According to transition probabilities of the system, HMM move from one state to another and produces sequence of observables from each state depending upon the emission probabilities.

IV. EXPERIMENT PROCEDURE

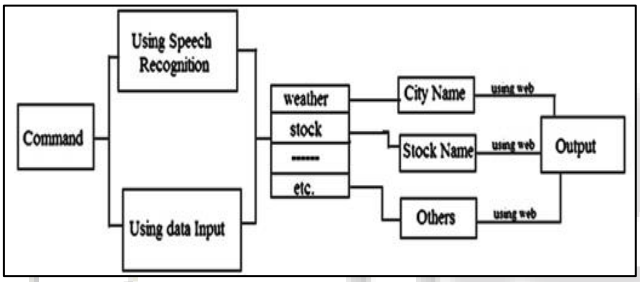

Fig. 1: Block Diagram of the system

First of all the user input is to ask to interest in knowing about stock or weather using speech recognition or using data input. The system includes two kinds of data i.e. Stock and Weather, In recognition of lesser number of words the Cross correlation is used for more efficiency.

In the next stage, we will use HMM for classification of the system involve multiple options for user to speak as HMM is more efficient for classification of large number of reference dataset. Once the word is recognized then data from the web extract and after that the data from web displayed and converted into speech using web based text.

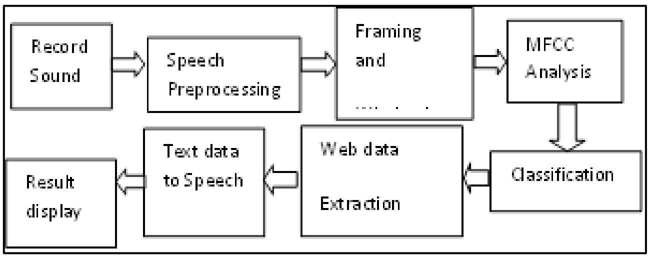

Fig. 2: Block Diagram Speech Recognition and Data Extraction

For training purpose a database of 40 users was created. We will record the speech file in .wav format using Audacity-win-2.0.6 sound recorder application for windows and save it in .wav format. The speech signal is recorded with the sampling rate of 16000 samples per second .Each user was asked to record 10 pronouncement for each of the items in stocks and cities list. We have used 10 names of different cities and stocks to generate a total of 400 samples/word.

Preprocessing involves techniques:

1) Pre-emphasis
2) Framing
3) Windowing.

As we know that the shape of vocal-cord is dynamic, different and rapidly changes with respect to time, so the generated signals will be divided into frames of small duration of time where the signals will be nearly static. The windowing technique is used in smoothening of edges of signal. The windowing function is used to avoid the sudden change in frequency response between adjacent frames. The hamming window is used to punctuate peaks of noise signals while suppressing spectral leakage of the signals. Each of these small duration frames will be used as input to the Cepstral analysis which is followed by the classification using Hidden-Markov-Model. During the testing, the maximum likelihood of the pronouncement of speaker is found and correspond word is recognized. Thus after the recognition of word the next task is to extract data information of stocks and weather from web server by using the API key of the website. And after that extracted data will be displayed and converted into sound using text to speech converter.

V. RESULT AND DISCUSSION

A. Cross-Correlation Method simulation result

Cross-correlation method is efficient for less number of words, the efficiency of the system is degrade if a large number of word-set is used. We have recorded reference speech signals for both stocks and weather information. Then we recorded target speech signals 10 times as:

First: 1 to 5 time for – Stocks information

Second: 6 to 10 time for – Weather information

These signals recorded at almost calm and noiseless environment using external microphone. The Cross-Correlation between the target signal and reference signal shown in Fig-3 (left side for Weather and right side for Stocks).

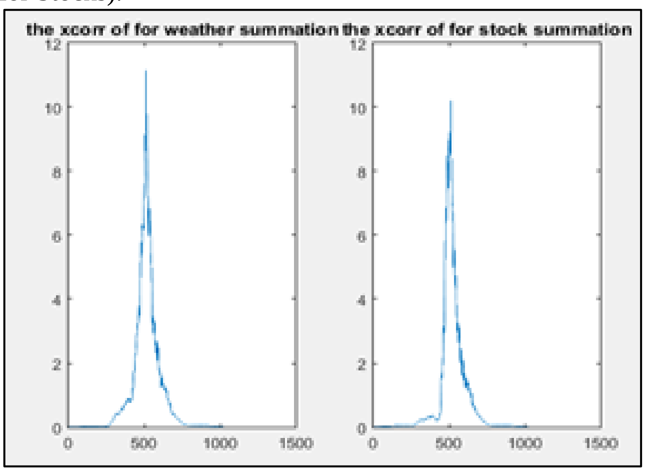

Fig. 3: Cross-correlations between the target signal and reference signals for Weather and Stocks

Test-Time	Frequency Shift (For Weather)	Frequency Shift (For Stock)	Final Result
1	0	21	Weather
2	0	11	Weather
3	-1	12	Weather
4	-2	10	Weather
5	0	8	Weather
6	3	-1	Stock
7	9	0	Stock
8	6	0	Stock
9	10	-1	Stock
10	9	-2	Stock

Table 1: Simulation Result for Frequency Shift

Table I. The frequency shift for each target signals of 'Weather' and 'Stocks' is shown in Table-I and the final result according to the frequency shift is produced. For two words in 10 iterations we observed 100% recognition efficiency and for 10 words it reduces to 65%. Thus we can say that for lesser number of words we will get good efficiency. So we observed that for lesser number of words this algorithm gives better efficiency and efficiency reduces with increase in the set of words.

B. *Hidden Markov Model simulation result*

Recognition efficiency of different techniques are given respectively for 10 set of words:
Linear Prediction Coefficient (LPC) – 66%
Linear Prediction Cepstral Coefficient (LPCC) – 72%
Mel-Frequency Cepstral Coefficient (MFCC) – 91.60%
The conclusion is that we have gained better efficiency by using MFCC technique, which is higher than LPC and LPCC techniques. The recognition efficiency results for different feature extraction techniques are shown in Fig. 4.

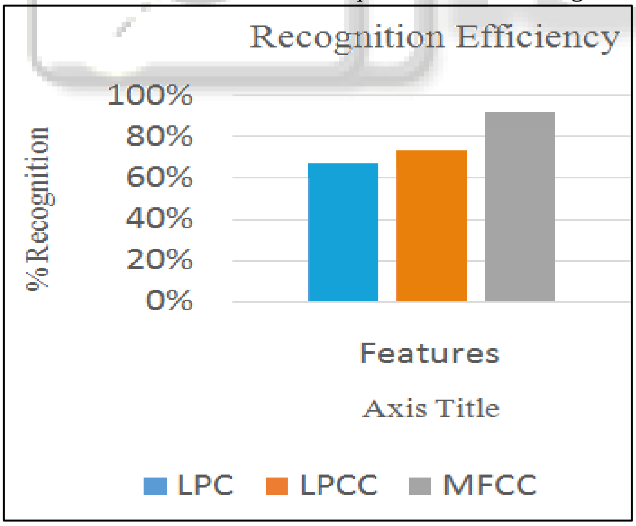

Fig. 4: Speech Recognition efficiency for different Feature Extraction Techniques

C. *Data extraction from web*

The likelihood word which is recognized by the system, after recognition it will looked onto the web server for the data of 'Weather' and 'Stocks'. The retrieved data will be displayed on the MATLAB console. Then the retrieved data will be displayed and the audible sound will be generated by using text to speech convertor.

VI. CONCLUSION

In this paper we have presented a scheme in which a person can extract data from web server just only by his/her voice by using speech recognition system. In Cross-Correlation Method we have gained the recognition efficiency 65% for the set of 10 words and for set of two words it was 100% , So it implies that the efficiency will degraded if the set of word increases. By using Hidden Marcov Process the recognition efficiency increase to 91.60% for the same 10 set of words which is 26.60% increases with respect to Cross-Correlation efficiency. The objective of this paper is to retrieve data information by using speech signal given by user's voice. We have successfully performed the extraction of data for 'Weather' and 'Stocks' by names using speech recognition system.

REFERENCES

[1] "Fundamental of Speech Recognition," L. R. Rabiner and B. H. Juang, I- edition, Pearson Education, Delhi, 2003.
[2] "Speech analysis and synthesis by linear Prediction of the Speech Wave," B. S. Atal and S. L. Hanauer, J. Acoust. Soc. AmVol. 50, No. 2, Aug. 1971, pp. 637-655.
[3] "Online Hierarchical Transformation of Hidden Markov Models for Speech Recognition " Jen-Tzung Chien, IEEE transactions on speech and audio processing, vol. 7, no. 6, november 1999.
[4] "On the Use of Windows for Harmonic Analysis with the Discrete Fourier transform", Fredric J. Harris , Proceedings of the IEEE, VOL 66, No.1, JANUARY, 1978.
[5] "Speech Recognition Based System to Control Electrical Appliances "Arvinder Singh, Gagandeep Singh, (IJEIT) Volume 2, Issue 2, August 2012.
[6] J. L. Flanagan, "Speech analysis and perception," Springer-Verlag, Berlin, 2nd edition, 1965.
[7] "Speech Recognition Using Correlation Technique " Anjalika Gupta1 , Prof. Pankaj Raibagkar2 , Dr. Anup Palsokar, (IJCTER) e-ISSN 2455–1392 Volume 3 Issue 6, June 2017 pp. 82 – 89.

