

SEARCH TIME REDUCTION USING HIDDEN MARKOV MODELS FOR ISOLATED DIGIT RECOGNITION

Sheena C V¹, T M Thasleema² and N K Narayanan³

Department of Information Technology, Kannur University, Kerala, INDIA
sheenacvg@gmail.com¹, thasnitml@hotmail.com² and
nk narayanan@gmail.com³

ABSTRACT

This paper reports a word modeling algorithm for the Malayalam isolated digit recognition to reduce the search time in the classification process. A recognition experiment is carried out for the 10 Malayalam digits using the Mel Frequency Cepstral Coefficients (MFCC) feature parameters and k - Nearest Neighbor (k-NN) classification algorithm. A word modeling schema using Hidden Markov Model (HMM) algorithm is developed. From the experimental result it is reported that we can reduce the search time for the classification process using the proposed algorithm in telephony application by a factor of 80% for the first digit recognition.

KEYWORDS

Isolated Digit Recognition, Mel Frequency Cepstral Coefficient, k - Nearest Neighbor, Hidden Markov Model.

1. INTRODUCTION

Speech recognition is one of the active research areas in Human Computer Interaction [1]. Speech Recognition is the ability of a computer to recognize general, naturally flowing utterances from a wide variety of speakers. It involves capturing and digitizing the sound waves, converting them to basic language units or phonemes, constructing words from phonemes, and contextually analyzing the words to ensure correct spelling for words that sound alike. This paper discusses two different stages for Malayalam digit recognition using Mel Frequency Cepstral Coefficients (MFCC) algorithm. In the first stage a recognition experiment is carried out using k-NN algorithm and in the later part a word modeling algorithm is proposed for the Malayalam telephony application using Hidden Markov Model (HMM) for faster classification.

The basic theory of HMM was introduced and studied in the late 1960s and early 1970s [2]. But from the literature study, it is reported that only in the past some decades only HMM has been applied accurately to problems in speech processing. These models are very rich in mathematical structure and thus it can provide the theoretical basis for use in wide range of applications. Here we have developed a word modeling algorithm for the Malayalam isolated digits using HMM.

Malayalam is the one of the major language in the Dravidian language family. It is regional language of south Indian state of Kerala and also on the Lakshadweep islands spoken by about 36 million people [3]. The phonemic structure of Malayalam contains 51 vowels/consonant-vowel sounds in which 15 long and short vowels and 36 consonant-vowel sounds. Due to lineage of Malayalam to both Sanskrit and Tamil, Malayalam language structure has the largest number of phonemic utterances among the Indian languages [4]. Malayalam script includes letters capable of representing all the phoneme of Sanskrit and all Dravidian languages [5]. In this work, in the recognition stage we have used ten Malayalam digits uttered by a single speaker repeated 20 times and is tabulated in the table 1. In the word modeling part we have used isolated digits in Malayalam for the use of telephone applications.

Table 1: Malayalam Digits

Digits	Words	English
0	പൂജ്യം	Poojyam
1	ഒന്ന്	Onnu
2	രണ്ട്	Randu
3	മൂന്ന്	Moonnu
4	നാല്	Nnalu
5	അഞ്ചു	Anju
6	ആറ്	Aaru
7	ഏഴ്	Ezhu
8	എട്ട്	Ettu
9	ഒമ്പത്	Onpathu

The organization of this paper is as follows. In section II feature extraction using MFCC algorithm is discussed. Section III explains recognition experiments using k-NN algorithm and the results are discussed. Section IV gives an overview on HMM followed by Malayalam isolated digit word modeling using the probability matrix. Finally in section V concludes the present work followed by directions for future work.

2. FEATURE EXTRACTION USING MFCC COEFFICIENTS

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. In this paper we discuss one of the basic speech feature extraction technique namely Mel Frequency Cepstral Coefficient (MFCC). The MFCC method uses the bank of filters scaled according to the Mel scale to smooth the spectrum and to perform similar to that executed by the human ear [6]. The filters with Mel scale spaced linearly at low frequencies up to 1 kHz and logarithmically at higher frequencies are used to capture the phonetical characteristics of the speech signals. Thus MFCCs are used to represent human speech perception models. MFCCs are computed as in fig 1.

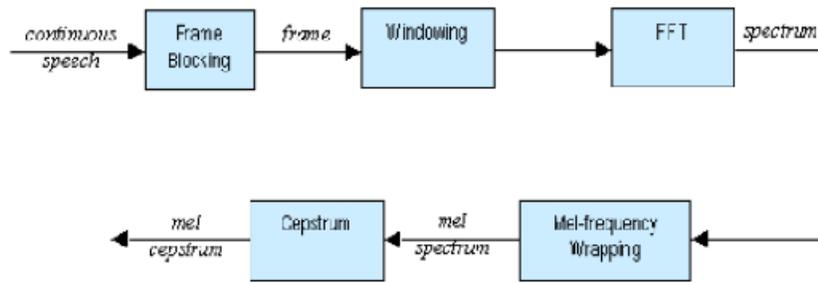


Figure 1: Block Diagram of MFCC Feature extraction process

Frame blocking is the process of segmenting the speech samples obtained from the analog to digital (A/D) conversion into small frames with time length in the range of (20 to 40) milliseconds. In the next step windowing is carried out to each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. After that Fast Fourier Transform (FFT) is applied for converting each frame of N samples from the time domain into the frequency domain. Then each frame with actual frequency, f measured in Hz is converted on a scale called the 'Mel' scale. The Mel frequency is calculated using the formula

$$F_{mel} = 2595 \log(1 + F_{Hz}) / 700 \quad (1)$$

The Mel-frequency scale is linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz. The log Mel spectrum is again converted into time domain using discrete cosine transform (DCT) to get Mel Frequency Cepstral Coefficients (MFCC). Thus the MFCC is derived by applying above described procedure for each speech frame. A set of MFCC coefficients are extracted by taking the average of each frame and are used as a feature set in the k-NN recognition algorithm.

3. SPEECH RECOGNITION USING K-NN ALGORITHM

k-Nearest Neighbor algorithm (k-NN) is part of supervised learning that has been used in many applications in the field of data mining, statistical pattern recognition and many others [7]. k-NN is a method for classifying objects based on closest training samples in the feature space. An object is classified by a majority vote of its neighbors. k is always a positive integer. The neighbors are taken from a set of objects for which the correct classification is known [8]. Hand proposed an effective trial and error approach for identifying the value of k that incurs highest recognition accuracy [9]. Various pattern recognition studies with highest performance accuracy are also reported based on these classification techniques [10].

k-NN assumes that the data is in a feature space. If $k=1$, then the algorithm is simply called the nearest neighbor algorithm. In the example in Fig. 2, we have three classes and the goal is to find a class label for the unknown example x_j . In this case we use the Euclidean distance and a value of $k=5$ neighbors. Of the 5 closest neighbors, 4 belong to w_1 and 1 belongs to w_3 , so x_j is assigned to w_1 , the predominant class.

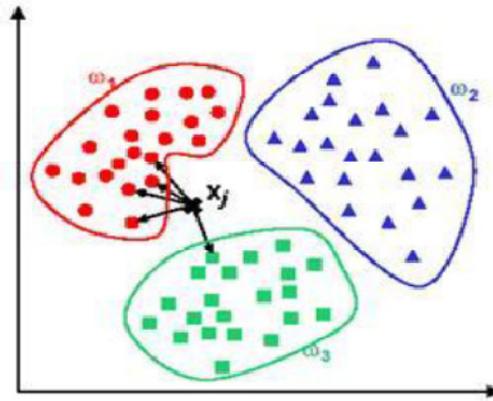


Figure 2. k-NN Classification Example

In the specified experiment a database of 20 repetitions of 10 Malayalam digits are used for testing and training purpose. One hundred samples are taken for training and one hundred samples for testing. An average recognition accuracy of 62% is obtained using k-NN algorithm for the Malayalam digit recognition.

4. WORD MODELLING USING HMM

Hidden Markov Model (HMM) is a statistical model in which the system being modeled is assumed to be a Markov process with unobserved state [11]. The model is completely defined by the set of parameters A , B and π where A is the transition probability, $A = \{a_{ij}\}, 1 \leq i \leq j \leq N$, B is the emission probability, $B = \{b_j(w_k)\}, 1 \leq j \leq N, 1 \leq k \leq M$ is the probability of the observation w_k being generated from the state j , π is the initial state probabilities. Thus a model of N state and M observation can be defined by $\lambda = (A, B, \pi)$ [2]. The present work discusses the modeling algorithm developed for the Malayalam isolated digit recognition in telephony application. We considered 50 mobile numbers of the BSNL service provider. Here we calculated and tabulated the initial probabilities and transition probabilities and are given in table 2 and the corresponding HMM model is shown in fig 3.

Table 2: Probability Matrix

Digits-(observation)	positions →									
	1	2	3	4	5	6	7	8	9	10
0	0	0	0.05	0.05	0.15	0.15	0.15	0.025	0.175	0.125
1	0	0	0	0	0.1	0.075	0.075	0.075	0.1	0.025
2	0	0	0	0	0.075	0.025	0.1	0.175	0.125	0.075
3	0	0	0	0	0.075	0.125	0.1	0.1	0.125	0.075
4	0	0.95	0.725	0	0.15	0.15	0.025	0.05	0	0.1
5	0	0.05	0	0.025	0.075	0.1	0.125	0.1	0	0.1
6	0	0	0	0.425	0.175	0.025	0.125	0.125	0.1	0.125
7	0	0	0	0.5	0.075	0.025	0.05	0.15	0.05	0.15
8	0.05	0	0	0	0.05	0.225	0.075	0.1	0.125	0.175
9	0.95	0	0.225	0	0.075	0.1	0.175	0.1	0.2	0.05

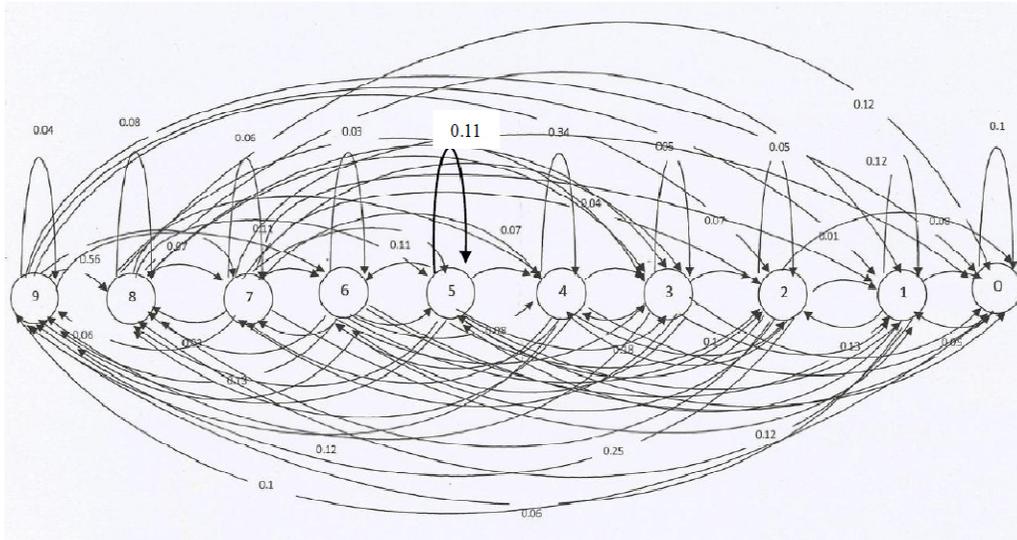


Figure 3. HMM Model

From the tabulated result it is found out that the initial probability for all the digits except the digit 8 and 9 is 0 while the digit 8 has the probability 0.05 and 9 has the probability 0.95, since the database contains all the BSNL numbers starting with the digit 8 and 9. In this work we make use this result to reduce the search time in the recognition experiment in such a way that in the classification stage we can start the recognition experiment only by considering the digit 8 and 9 and hence we can reduce the search time by a factor of 80% in the first digit. The similar procedure can be extended to the successive digits also, resulting in a good reduction in search time in recognition/classification experiment.

5. CONCLUSIONS

This paper presented a word modeling schema for the recognition of Malayalam isolated digit recognition using various mobile numbers uttered in Malayalam. Two different stages are carried out in this study. In the first stage a recognition experiment is carried out using MFCC coefficients and k-NN algorithm for the 10 Malayalam digits and a recognition accuracy of 62% is obtained. In the second stage a word modeling algorithm is proposed for the Malayalam isolated digits using HMM. From the experimental result it is reported that using proposed algorithm we can reduce the search time by a factor of 80% in the recognition of first digit for the classification process of BSNL telephone number recognition system. The modeling of the all the isolated digits from the different service providers using HMM modeling algorithm and its recognition using other classification algorithms are some of our future research directions.

REFERENCES

- [1] Rabiner Lawrence and Biing-Hwang Juang (1993) Fundamentals of speech Recognition Prentice Hall.
- [2] L. R. Rabiner and B. H. Juang, (1986), "An Introduction to Hidden Markov Models", IEEE ASSP Magazine, pp. 4 – 16, 1986.
- [3] Ramachandran, H. P (2008) Encyclopedia of language and linguistics, Oxford: Pergamon Press.
- [4] Aiyar, S (1987). Dravidian theories, p. 286.

- [5] Govindaraju, V., & Setlur, S (2009), Advances in pattern recognition. Guide to OCR for Indic scripts: Document recognition and retrieval, Berlin: Springer. (p. 126).
- [6] Ibrahim Patel and Y Srinivas Rao, (2010), “ Speech Recognition using HMM with MFCC analysis using frequency spectral decomposition technique”, Signal and Image Processing - An International Journal, Vol. 1(2), pp. 101 – 110.
- [7] Zhang. B and Srihari S N, (2004), “Fast k – Nearest Neighbor using Cluster Based Trees”, IEEE trans. on Pattern Analysis and Machine Intelligence, Vol. 26(4), pp. 525 – 528.
- [8] Pernkopf.F,(2005),“Bayesian Network Classifiers versus selective k–NNClassifier Patter Recognition, Vol. 38, pp. 1 – 10.
- [9] Hand D J (1981) Discrimination and classification, New York, Wiley.
- [10] Ray A. K and Chatterjee B, (1984), “Design of a Nearest Neighbor Classifier System for Bengali Character Recognition”, Journal of Inst. Elec. Telecom. Eng , Vol. 30, pp 226 – 229,.
- [11] Daniel Jurafsky and James Martin (2004) Speech and Language Processing, Pearson Education.

Authors

Sheena C V received her MSc in Computer Science from, Kannur University, Kerala, India in 2008, she is currently a Ph.D. student under Prof Dr.N.K.Narayanan at Department of Information Technology, Kannur University, Kerala, India. Her research interests include Computer Vision, Digital Image Processing, Digital Speech Processing, Artificial Intelligence and Artificial Neural Networks.



T M Thasleema had her M Sc in Computer Science from Kannur University, Kerala, India in 2004. She had to her credit one book chapter and many research publications in national and international levels in the area of speech processing and pattern recognition. Currently she is doing her Ph.D in speech signal processing at Department of Information Technology, Kannur University under the supervision of Prof Dr N. K Narayanan.



Dr. N.K. Narayanan is a Senior Professor of Information Technology, Kannur University, Kerala, India. He earned a Ph.D in speech signal processing from Department of Electronics, CUSAT, Kerala, India in 1990. He has published more than hundred of research papers in national & international journals in the area of Speech processing, Image processing, Neural networks, ANC and Bioinformatics. He has served as Chairman of the School of Information Science & Technology, Kannur University during 2003 to 2008, and as Principal, Coop Engineering College, Vadakara, Kerala, India during 2009-10. Currently he is the Director, UGC IQAC, Kannur University.

