A Multi-HMM Marathi Isolated Word Recognizer

Kayte Charansing Nathoosing

Department Of Computer Science, Indraraj College Sillod, 431112, Dist Aurangabad, (M. S.) India
charankayte@gmail.com

ABSTRACT

Punjabi, Hindi, Marathi, Gujarati, Sindhi, Bengali,Nepali, Sinhala, Oriya, Assamese, Urdu are prominent members of the family of Indo-Aryan languages. These languages are mainly spoken in India, Pakistan, Bangladesh, Nepal, Sri Lanka and Maldive Islands. All these languages contain huge diversity of phonetic content. In the last two decades, few researchers have worked for the development of Automatic Speech Recognition Systems for most of these languages in such a way that development of this technology can reach at par with the research work which has been done and is being done for the different languages in the rest of the world. A Multi-HMM speaker – independent Marathi isolated word recognition system is described in this system. Three vector quantization methods, the LBG, EM, and a new MGC algorithm, are used for the classification of the speech space. This multi-HMM system result in an improvement of about 50 percent in the error rate in comparison to the single model system

Keywords: Marathi, Isolated Word Recognizer.

INTRODUCTION

Speech recognition to allow the machine to catch words phrases and sentences that we speak (Tarun Pruthi, 1993). Currently, the most popular approach to speech recognition is the combination of Vector Quantization (VQ) with a hidden Markov Modeling (HMM). To divide the single space into a number of cells or sub-space and produce a codebook of vector (Yaxin Zhang, 1992). The second step, Hidden Markov Modeling, is used to produce a set of model which represents possible sequence of codebook vectors which arise from words that the system is to recognize. The commonest VQ algorithm is LBG algorithm. It has the advantage of being simple and not requiring excessive computation. The LGB algorithm does not guarantee that the classification of speech is globally optimal (Desilva, 1994). This means that some of the codebook vectors may not be typical of the vector in the cell they represent. The shortcoming of the LBG algorithm lead to an inappropriate classification of the speech space and inadequate matching with the Hidden Marko Modelling and consequently limited recognition accuracy for the whole system.

Observation has shown that different utterance of the same speech sound form a cluster around some center, which represents some average or fiducially production of the sound. The variation about the mean will occur at random when a large population of speaker is considered, so the points in the cluster may be distributed according to multidimensional Gaussian probability density function (Yaxin Zhang, 1992). This view of the speech production process suggest that classification of the speech space is bettor done on the base of a Gaussian Mixture Model(GMM),in which the point are clustered around the means according to Gaussian distributions and each cluster is assigned a weight representing the frequency with which point in the cluster occur. A method now known as Expectation-Maximization (EM) algorithm for estimating the parameters of GMMs was described by Wolfe (Dempster, 1977). The EM algorithm can be used as a substitute for the LBG algorithm for quantization of the speech space. Experiments have show that the EM algorithm matches the HMM quite well and leads to a better recognition accuracy. However, the EM algorithm is more computation intensive than the LBG algorithm and is sensitive to background noise. Another classification method that we have devised is the MGC algorithm which is similar to the EM algorithm but requires less computation in training and produces slightly rougher classification than the EM algorithm.

In this paper we describe a multi-HMM(MHHM) speaker –independent Marathi isolated word recognizer In which the three VQ algorithm mentioned above are used independently of each other. Theses quantization's of the speech space are then used to produce three HMMs for each word in the vocabulary using the Baum-welch algorithm. In the recognition step, the Viterbi algorithm is used in three sub-recognizers. The probabilities of the observation sequence matching the method are multiplied by the weights and summed to give the probability that the utterance is of a particular word in the vocabulary. We report the result of comparing this method with the use of a single vector quantization algorithm. This result is a reduction of about 50 percent in the error rate in comparison to the best single VQ/HMM system.

MATERIALS AND METHODS

Author constructed a multi-Hidden Markov Model (MHHM) speaker-independent Marathi isolated word recognizer. The system is composed of three sub-recognizer, each of which we used one vector quantization method for the first step modeling.

In this training step, the three classification algorithm mentioned above are employed for a parallel vector quantization and a codebook is generated for every sub-recognizer .Thus ,three models for each word of the vocabulary are produced by the Hidden Markov Modeling (Baum-Welch) algorithm. In the recognition step, the Viterbi algorithm is used in parallel with the three sub-recognizers (Yaxin Zhang, 1992). The probabilities of the Observation Sequences matching the models are multiplied by weights determined by the individual recognition accuracies. The weighted probabilities for each word to be recognized are summed .Then the model which produce the highest probability is the output.

Table 1 demonstrates the recognition improvement of the MHMM/VQ system. The Marathi numbers in the table were the outputs scores of individual HMM/VQ and MHMM/VQ recognizers. The highest score is the system output the whole system gave a correct output even when two of three single systems, LBG and MKF, give incorrect recognitions.

Table 1: The scores of Hidden Markov Models matching the input word "ek" with three vector quantization algorithms.

	HMM/LBG	HMM/MKF	HMM/EM	MHMM/VQ
0 ("shoonya")	-177	-173	-177	-791
1 ("ek")	-134	-122	**-111**	**-539**
2 ("do")	-191	-188	-184	-841
3 ("teen")	-173	-165	-169	-759
4 ("char")	-156	-148	-151	-680
5 ("paanch")	-166	-169	-163	-746
6 ("saaha")	-189	-197	-200	-885
7 ("saat")	-178	-202	-191	-840
8 ("aath")	-195	-186	-183	-840
9 ("nau")	**-129**	-131	-140	-606
OH	-141	**-119**	-144	-608
OUTPUT	9 ("nau")	OH	1 ("ek")	1 ("ek")

RESULTS AND DISCUSSION

A set of evaluation test was performed on the MHMM/VQ system. The training data comprised a small vocabulary of eleven Marathi isolated digit (from zero to nine and OH) spoken by 20 speakers (10 male and 10 Female) and tested data spoken by 21 different speakers (10 male and 11 female). For preparing the input for the VQ system, the speech data was windowed and feature vectors were constructed for each window (Linde, 1980). The first pre-processing step was the computation of the power spectrum of the windows signal using a FFT routine, followed by summation of the components of the power spectrum to simulate a bank of twelve bands –pass filters.

A set of preliminary investigation was performed using the individual HMM/VQ system. Different codebook size where used and table 2 show the results.

Table 2 showed among the individual recognizer the HMM/EM gave the best recognition accuracies, and the HMM/LBG the worst. This is consistent with our expectation that the Gaussion mixture model is a good description of speech feature which matches Hidden Markov Modeling very well (Rabiner, 2000).

Table 3 showed the result achieved by multi-model system. Compared with the best single recognizer, HMM/EM, the MHMM/VQ system obtained 38.8%, 54.9%, and 53.1% reeducation In the recognition error rates for the codebook size 32, 64 and 128 respectively.

Table 2: Individual HMM/VQ recognition

Code Size		32	64	128
Accuracy of recognition	LBG	86.18%	93.09%	94.33%
	MKF	87.78%	93.13%	95.34%
	EM	91.05%	93.59%	97.23%

Table 3: Multi HMM/VQ recognizer

Codebook Size	32	64	128
Accuracy of recognition	94.53%	97.11%	98.70%

Consideration of the speech production process suggests that Gaussian Mixture Models offers a good description and the EM algorithm is an effective classification method for the first step modeling in a HHM/VQ system (Dempster, 1977).

The multi-Hidden Markov Model speech recognition gave much better recognition results (Yaxin Zhang, 1992). This was proved by the performance of combinations of any two of three sub recognizers and combination of three of them together. The best result achieved by MHMM/VQ recognizer represent a reduction in the even rate of about 50 percent in comparison to the HMM/EM recognizer, in the best single recognizer.

LITERATURE CITED

Dempster AP, Laired NM and Rubin DB, 1977. Maximum Likelihood for Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological),* **39**(1):1-38.

Desilva CJS, Togneri A, Alder M Attikiouzel, 1994. Speaker-independent isolated word recognition using multiple hidden Markov models, *Institute of Electrical and Electronics Engineers*, **141**: 197-202

Linde Y, Buzo A and Gray RM, 1980. An algorithm for vector Quantizer Design, *Institute of Electrical and Electronics Engineers Transaction on Communication.* **28**(1):84.

Rabiner LR, **2000.** A tutorial on Hidden Markov Models and selected Applications in speech Recognation, *Proceedings of Institute of Electrical and Electronics Engineers,***77**(20): 257-286.

Tarun Pruthi, Sameer Saksena, Pradip K Das, 1993. Swaranjali: Isolated Word Recognition for Hindi Language using VQ and HMM. *International Journal of Computing and Business Research,* **2**(2):3.

Yaxin Zhang, Christopher JS, deSilva, Roberto Togneri, Mike Alder and Yianni Attikiouzel, 1992. A Multi-HMM isolated word recognizer**,** *The Australasian Speech Science and Technology Association (ASSTA), SST-***92**:568-573.

Wolf JH, 1970. Pattern Clustering by multivariate Mixture Analysis. *Multivariate Behavioural Research,* **6**:329-350.