International Conference on
Information, Communications and Signal Processing
ICICS '97
Singapore, 9-12 September 1997

Scalar Quantization of Features in Discrete Hidden Markov Models

Sitaram Ramachandrula and Sreenivas Thippur

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore-560012, India
email: sitaram@protocol.ece.iisc.ernet.in & sreeni@iis.fhg.de

Abstract

Traditionally, discrete hidden Markov models (D-HMM) use vector quantized speech feature vectors. In this paper, we propose scalar quantization of each element of the speech feature vector in the D-HMM formulation. The alteration required in the D-HMM algorithms for this modification is discussed here. Later, a comparison is made between the performance of D-HMM based speech recognizers using scalar and vector quantization of speech features respectively. A speaker independent TIMIT vowel classification experiment is chosen for this task. It is observed that the scalar quantization of features enhances the vowel classification accuracy by 8 to 9 %, compared to VQ based D-HMM. Also, the number of HMM parameters to estimate from a given amount of training data has drastically reduced in the new idea.

1 Introduction

Currently the most successful speech recognition systems are based on hidden Markov models (HMM). A HMM provides a statistical representation of a highly variable speech signal. HMM models the sequence of speech feature vectors as a piecewise stationary process. It consists of a finite number of states, where each HMM state characterizes a stationary speech segment by using an observation probability density function.

There are two versions of HMM formulations. One in which the speech feature vectors are characterized by using a mixture of multi-variate probability density functions within each state, this is referred to as continuous HMM (C-HMM). In the second version of HMM, first the sequence of speech feature vectors are vector quantized using the finite sized codebook and then the corresponding sequence of VQ codeword indices are modelled by HMM by using discrete probability distributions (spanning over the codewords) within each state, this is generally referred to as discrete HMM (D-HMM). In this paper the VQ based D-HMM is referred to as VQ-D-HMM. All the algorithms of these two versions of HMMs required for speech recognition, have already been derived [1, 2]

The VQ-D-HMM has the problem of an amount of vector quantization error associated with it, compared to C-HMM, where the unquantized feature vector is directly used in the probability evaluation. But in C-HMM, the probability evaluation of observation vector is computationally intensive compared to VQ-D-HMM where it is just a table lookup, which is helpful in real time implementation of speech recognition systems. Based on the requirement and specifications of a speech recognition task at hand, the discrete or continuous HMM is chosen.

Normally the research in HMM aims at increasing the speech recognition accuracy by using ideas like incorporating speech knowledge in HMMs and using better speech features, etc. But there is also a requirement of new ideas in HMMs which can reduce the quantization error in VQ-D-HMM (may not be as good as C-HMM), but are computationally not expensive like C-HMM, and have less number of parameters to estimate [3]. This can result in better recognition accuracy compared to VQ-D-HMM. In this paper, one such HMM formulation based on scalar quantization of each element of the feature vector is proposed, this is called as Scalar-HMM. The Scalar-HMM formulation is very similar to the standard VQ-D-HMM [1].

Section 2 explains the Scalar-HMM and alterations required for it in the algorithms of VQ-D-HMM. An experimental evaluation of Scalar-HMM on a TIMIT

0-7803-3676-3/97/$10.00

vowel classification task is given in Sec. 3, followed by discussions in Sec. 4 and conclusions in Sec. 5.

2 Scalar HMM

There is not much difference in theory between VQ-D-HMM and Scalar-HMM except for the way the features are quantized. In Scalar-HMM, each element in the observation feature vector (e.g., each cepstral coefficient) is scalar quantized using a separate scalar codebook, unlike the vector quantization of the full vector done in VQ-D-HMM. In Scalar-HMM formulation, each HMM state will have a number of observation probability distributions which are equal to the number of elements in the feature vector, i.e., one probability distribution for each feature element (spanning over corresponding scalar codes). Except for this difference and its effect on the algorithms, there are no differences between Scalar-HMM and VQ-D-HMM.

The main effect of Scalar-HMM formulation on the three standard algorithms of HMM is the change in the way the probability of a given observation feature vector is calculated. In VQ-D-HMM, it is the direct table lookup after VQ, but in Scalar-HMM, to find the probability $P(\vec{O})$, of an observation vector $\vec{O}$, from a given HMM state j, first, each element in the given vector is scalar quantized using respective scalar codebooks. Then the product of probabilities of scalar codes (corresponding to each quantized feature element) from respective probability distributions from the state j gives the desired probability. This is explained below:

Let $\vec{O} = (x_1, x_2, ..., x_D)$ be the given observation vector with dimension D. After quantization with respective scalar codebooks, it is represented as $\vec{Y} = (y_1, y_2, ...y_D)$. Let $b_{jm}(k_m)$ be the (scalar) observation probability distribution corresponding to feature element m, $(1 \leq m \leq D)$, in state j, with k_m varying over the set of possible symbols N_m of feature element m, i.e., $1 \leq k_m \leq N_m$. Now the probability of the given observation vector $\vec{O}$, from state j is:

$$P_j(\vec{O}) = b_{j1}(y_1)b_{j2}(y_2)...b_{jD}(y_D) = \prod_{i=1}^{D} b_{ji}(y_i) \quad (1)$$

Here there is an assumption that the feature elements $x_1, x_2, ...x_D$, are independent of each other. It is observed that this methodology performs well in the case of cepstral features. This method provides a natural way of adding any new feature element to the feature vector without having to normalise it by its variance, which is required in VQ-D-HMM.

3 Experimental Evaluation

3.1 Task and Database

The idea of Scalar-HMM is tested on a speaker independent TIMIT vowel classification task. The vowel classification task is chosen because vowels are discriminated mainly by their spectral content, and the scalar vs. vector quantization considered in this paper addresses mainly the modelling of the spectral property and not the temporal structure of the sound. Eleven vowels $\{aa, ae, ah, ao, eh, ih, ow, uh, uw, ux, iy\}$ are sliced out from the training and test set of dialect-2 sentences of TIMIT acoustic phonetic database. TIMIT contains continuous speech sentences which are sampled at 16 KHz and are labelled into phonemes. The TIMIT database is divided into 8 divisions based on 8 dialects in America. In this experiment, speech sentences from dialect 2 are used. There were totally 76 speakers (53 male, 23 female) in training set and 26 speakers (18 male, 8 female) in test set of dialect 2. From each speaker 5 occurrences of every vowel is sliced out, thus there are 380 occurrences of each vowel in the training set and 130 occurrences in the test set.

3.2 Pre-Processing

The entire speech database created as explained above is analysed in frames of 15 ms with an overlap of 5 ms between frames. The pre-emphasis factor used is 0.95. From each frame of speech, 18 LPC derived cepstral coefficients are determined.

3.3 Baseline Recognition System

Before proceeding with the implementation of Scalar-HMM based vowel recognizer, for comparison the recognizer is also implemented using VQ-D-HMM. The vector quantization codebook required in this experiment is designed using LBG algorithm [4]. Later feature vectors of entire vowel database, are vector quantized using the designed codebook. Now using the VQ training sequences of each vowel, three state left-right HMMs are trained using Baum-Welch algorithm. The recognition results of the test-set and the training-set using these HMMs are reported in Tab. 1. The experiment is repeated for various sizes

Table 1: Vowel Classification Accuracy using VQ-D-HMM

Size of codebook	Train Set	Test Set
128	50.73%	40.57%
256	62.18%	41.47%
512	73.91%	43.47%
1024	86.29%	41.87%

Table 2: Vowel Classification Accuracy using Scalar-HMM

No. of codebooks	Size of codebooks	Train Set	Test Set
18	8	55.19%	50.34%
18	16	60.17%	50.45%
18	32	65.35%	49.95%

of VQ codebooks. It can be seen as the VQ codebook size increases, at some stage, the performance decreases, as the number of HMM parameters to estimate have become very large and the training data available is fixed.

3.4 Vowel Recognition using Scalar-HMM

The vowel recognition experiment is repeated using the new Scalar-HMM formulation with the same HMM specifications, i.e., the number of states etc., and same feature specifications. The only difference being the usage of scalar quantization of feature elements. Here first, the scalar quantization levels (i.e., scalar quantization codebooks) are designed using the LBG algorithm [4] for each of the feature element. Then the elements of all the feature vectors, of entire vowel database, are scalar quantized using the designed scalar codebooks. Later the vowel Scalar-HMMs are trained using the scalar quantized training sequences. The vowel recognition results of the test and training set using scalar-HMMs are given in Tab. 2. This experiment is done for various sizes of scalar codebooks.

4 Results and Discussions

It can been seen from Tab. 2 that the results of Scalar-HMM based vowel recognition have shown a remarkable improvement of 7 to 9 % on test set, compared to VQ-D-HMM results reported in Tab. 1. Most importantly the number of scalar codewords used (required) for each feature element in scalar HMM is much less than the size of VQ codebook used in VQ-D-HMM. This helps in faster codebook search and secondly, the number of parameters (observation symbol probabilities) to estimate have gone down considerably for a given amount of repetitions of the training data pertaining to each parameter.

The scalar or vector quantization of speech feature vectors in HMMs is only a representation of speech for facilitating speech modelling using HMMs. Representation of speech in any scenario (speech coding, speech recognition etc) aims to be accurate, with less difference (error) between the original and the representation. But a trade off is normally reached on the degree of accuracy wanted based on the available resources and the end application. In speech recognition using HMMs there are some factors which have to be considered before thinking of high accurate representation. If in VQ-D-HMM the VQ codebook size is increased so as to reduce the VQ error, then the number of output symbol probabilities to estimate (with the available training data) will rise, causing problems in robust parameter estimation, thus resulting in poor recognition accuracy. Therefore the size of VQ codebook is fixed based on the allowable accuracy of representation of speech, for which the number of HMM parameters to estimate are feasible for the available training data, such that the speech recognition accuracy achieved is the best. In speech compression scenario the VQ of speech features may give a more accurate representation of speech than the scalar quantization for comparable number of bits used. But in speech recognition using HMMs, with the available training data, the trade-off reached in Scalar-HMM, on the number of scalar quantization levels for each feature, i.e., number of HMM parameters to estimate (such that the recognition accuracy is the best) resulted in better speech recognition accuracy than the trade-off reached similarly in VQ-D-HMM case. The Scalar-HMM resulted in higher recognition accuracy than VQ-D-HMM for the optimum number of codewords fixed in each case considering all the factors. That means, for the number of levels fixed based on the factors discussed above, the scalar quantization of features represents speech more accurately than VQ though in the limiting case this may not be true.

The performance of the state-of-the-art vowel classifiers is higher compared to our classifier, as we used only cepstral coefficients as the features and did not use any other features like differential cepstral co-

efficients etc. The duration, language and context modelling etc, which are known for their contribution in performance improvement were not used, as the main idea here is to compare the performance of vector vs. scalar quantization in discrete HMM. The results show that the Scalar-HMM is able to recognize unseen speaker data very well compared to VQ-D-HMM.

This work is a step ahead of the multi-codebook-HMM idea of Gupta et al [5], where they use different VQ codebooks for different sub-vectors constituting the main feature vector, e.g., they have used two codebooks, one for cepstral vector and one for differential cepstral vector. One such experiment is done here by using two sub-vectors each of dimension 9 (with 18 being the dimension of the main feature vector), the first 9 cepstral coefficients belong to one sub-vector and the next 9 coefficients belong to the second sub-vector. Each sub-vector was then vector quantized using separate codebooks, having 128 codewords each. The accuracy of the vowel recognizer using this multi-codebook-HMMs is 66.62 % and 46.26 % for training and test set respectively. We have conducted many experiments by splitting the main feature vector into different number of sub-vectors. The recognition results of all these experiments on test data are better than the standard VQ-D-HMM but always inferior than the scalar-HMM.

Apart from recognition performance improvement, some other advantages of the Scalar-HMM are: for a given amount of training data, the number of quantization levels to be estimated and thus the number of HMM parameters to be estimated is less, which helps in estimating robust parameters. Thus the ratio of available training data to number of parameters to estimate, has increased in Scalar-DHMM.

5 Conclusions

In this paper the scalar quantization of each element of the speech feature vector, in the discrete HMM formulation is proposed. A comparison is made between the standard VQ based discrete HMM and the Scalar-HMM proposed here. In a speaker independent vowel recognition experiment, it was shown that the Scalar-HMM resulted in higher recognition accuracy compared to VQ-D-HMM.

References

[1] Rabiner, L. R., "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of IEEE, vol. 77, no. 2, pp. 257-285, Feb. 1989

[2] Liporace, L. A., " Maximum likelihood estimation for multi-variate observation of Markov sources", IEEE Trans. Information theory, Vol. IT-28, No. 5, pp 729 - 734, 1982.

[3] S. Sagayama and S. Takahashi., "On the use of scalar quantization for fast HMM computation", Proc. ICASSP 1995, pp. 213-216.

[4] Y. Linde, A. Buzo and R.M. Gray, "An algorithm for vector quantizer design, IEEE Trans. Communication, Vol. COM-28, No. 1, pp. 84-95, Jan. 1980.

[5] Gupta, V. N., Lennig, M., and Mermelstein, P.,"Integration of acoustic information in a large vocabulary word recognizer", Proc. IEEE ICASSP'87, pp. 697-700, 1987