REVIEW OF MULTI-GROUP MIXTURE WEIGHT HMM

Kaihua Huang

Department of Electrical and Computer Engineering
Mississippi State University
Mississippi State, MS 39762 USA
email: huang@isip.mstate.edu

ABSTRACT

Hidden Markov Models are stochastic models capable of statistical learning and classification. They have been widely applied in most of state-of-the-art speech recognizer because of their great adaptability and versatility in handing sequential signals. Normally, more Gaussian components were used for each state, less error rate the recognizer will produced but at the same time it will increase the computation significantly. Additionally, since the weight of each Gaussian component trained from traditional method only depicts the average of Gaussian component probabilities for the training data, they are not appropriate to be used to classify some particular kind of speech signal.

A new method called Multi-group Mixture Weight HMM (MGMW HMM) is proposed by L. Ming and Y.T.Cheng to solve this problem with very little additional computation. Also, the authors claim that they achieved 12% error reduction compared to the traditional continuous HMM approach. This paper focuses on analysis the underlying theory of MGMW-HMM method. The derivation of the MGMW formula is investigated. The effectiveness of this approach is evaluated. A conclusion of objection for the method is made based on detailed analysis of this method.

1. INTRODUCTION

In speech recognition, the underlying assumption of the HMM is that the speech signal can be well characterized as a parametric random process, and that the parameters of the stochastic process can be estimated in a precise, well-defined way. In mixture density HMM, the modeling of the stochastic observation processes associated with the states is based on the estimation of the probability density function of the short-time observations in each state as a mixture of Gaussian densities.As we know, an HMM can be completely characterized by a matrix of state transition probability, observation densities, and initial state probability. Previous study of these probabilities has showed that the observation densities are most important for the performance of those recognizers using HMMs. Most of the improvement of HMM is made in this respect. As mentioned above, we want to improve the performance of recognizer, a straight forward way is to use more Gaussian density functions for each state. Because of large amount of logarithmic and exponential computation, it will take much more time for training and recognition than the HMMs with fewer Gaussian density functions. MGMW HMM is such a method claimed to resolve this kind of deadlock.

Equations (2) and (4) show that the weight of each component is the statistical average of the component probabilities. Accordingly, these weights can be used to describe the average characteristics of the corresponding state. Usually, some components have high weight contribute significantly while those with low weights contribute little. Therefore, the characteristics of these states are characterized mainly by those components which have high weights. But in some cases, the characteristics of speech are closer to that of those components which have low weights. Thus, these speech, however not commonly, can't be recognized properly. The proposed method set each state has several groups of component weights such that it can meet different cases. But why it can meet different cases is not state clearly, see section 4.

In the next section we briefly review the model of

MGMW-HMM, and outline the key equations in particular. Section 3 then shows some experiments conducted by the authors. Section 4 presents analysis of the experiment results and some doubts are extended on this method.

2. MGMW HMM

Obviously, for any model using HMM, the most difficult problem is to determine a method to adjust the model parameters. The MGMW HMM is trained by using the Baum-Welch method (also known as the EM (expectation-maximation) method) to get parameters.Because of the importance of initial parameters for continuous density HMM, the first issue is to get proper initial component weights. Initially by using this approach, as mentioned by the authors, for each frame in an speech signal utterance a vector called mixture weight vector is calculated and stored in the vector pool of the corresponding state. Then all mixture vectors in the vector pool of one state are classified into several clusters with center vector of each cluster corresponding to a group of initial mixture weight. After this procedure, MGMW HMM can learn through EM algorithm, which, of course, involves every group of mixture weight. For speech recognition, an appropriate group of mixture weight is selected by calculation and comparison of different output probabilities.

2.1. Retrieval of Initial MGMW

To get proper initial parameters of multi-group mixture weights for HMMs. Equ.(1), a widely used equation in continuous density HMM, is introduced.In (1), $\xi_t(j, k)$ is the probability of being in state j with the kth mixture components at time t, given the model λ and observation X.

$$\xi_t(j, k) = f(s_t = j, k_t = k | X, \lambda) \tag{1}$$

For the observation vector X_t at time t and given the model λ, the following variables are defined,

$$\eta_t(j) = \sum_{k=1}^{M} \xi_t(j, k) \tag{2}$$

$$F_j(X_t) = \left(\frac{\xi_t(j, 1)}{\eta_t(j)}, \ldots, \frac{\xi_t(j, M)}{\eta_t(j)} \right) \tag{3}$$

where, M is the number of mixture components, $F_j(X_t)$ is a M-dimension vector, called mixture weight vector.

To obtain the initial parameters of continuous density HMM, this MGMW HMM needs some speech data which has been labeled by hand or by force Viterbi decoding. After $F_j(X_t)$ is calculated, all mixture weight vectors $F_j(X_t)$ for the state j of the model λ are put together and classified by using classical clustering algorithm. The center vector of each cluster represents one group of initial mixture weights.

In some cases, some of the initial weights have a high value nearly to 1 whereas some of them are almost zero. Considering the performance of the continuous density HMM. The authors suggest to solve this problem by smoothing the mixture weights as showed below.

A group of mixture weight $F = \{c_1, c_2, \ldots, c_M\}$ can be smoothed by Equ.(4),

$$c_i' = c_i \cdot \theta + (1 - c_i) \cdot (1 - \theta) \tag{4}$$

where,

$$\theta = \frac{\varepsilon \cdot (M - 1)}{1 + \varepsilon \cdot (M - 2)} \qquad \varepsilon = 0.5 \sim 0.7 \tag{5}$$

and $1 \le i \le M$.

After this transformation, c_i' should be normalized by their sum.

Let,

$$C' = \sum_i c_i' \tag{6}$$

$$c_i'' = c_i' / C' \qquad 1 \le i \le M \tag{7}$$

Then the group of initial mixture weight will be

$$F = \{c_1'', c_2'', \ldots, c_M''\}.$$

2.2. Training of MGMW HMM

For each observation vector X_t, which is assigned to the state j of the model λ, the mixture weight vector and the output probability for each group Each group can be calculated. The mixture weight vector which gives the maximum output probability and the index of the mixture weight group is noted down. After each iteration, all mixture weight vectors of pth group of mixture weight of state j, model λ are added up and then divided by the number of the vectors which gives us the mixture weight vector for the next iteration. The re-estimate formula for the mixture weight can be written as (8).

$$\hat{F}_{p,j,\lambda} = \frac{1}{T}\sum_{t=1}^{T} F_{p,j,\lambda}(t) \tag{8}$$

where, $F_{p,j,\lambda}(t)$ is the mixture weight vector at time t which is assigned to the pth group of mixture weight, state j and model λ after Viterbi decoding [9]. After EM iterations, model parameters can be learned.

2.3. Recognition with MGMW HMM

For the recognition, there is no much difference between the traditional HMM and multi-group mixture weight HMM. This approach simply chooses the group of mixture weight which generate the maximum output probability for every observation as equation (9).

$$b_j(o_t) = \sum_{k=1}^{M} c_{p,k} \cdot N(o_t,\mu,\Sigma) \tag{9}$$

where, $b_j(o_t)$ is the output probability for observation o_t at state j, $c_{p,k}$ is the kth component weight of the pth group, $N(\ldots)$ is the Gaussian density function.

To reduce the increase of computation for MGMW HMM, when calculating the output probabilities, the probabilities of the Gaussian density functions can be calculated first, then they are multiplied with each group of mixture weight respectively.Hence, only several extra multiplication operations for each state are need and compared with the exponential operations, this additional computation can be ignored.

3. EXPERIMENTS

An experiment is made on a digital string recognition system by using MGME HMM method. The training data include 40 persons' speech data and the test data include 6 persons'. Each person has 50 utterances of digital strings. In order to compare with the traditional continuous density HMM, some experiments using traditional HMM were also made. In MGMW HMM method, three groups of mixture weight for every state are used. modeling methods.

Table 1: Experiment results for the performance of the traditional HMM and Multi-group Mixture Weight HMM

	Traditional HMM	MGMW HMM
Digital Correct	97.68%	98.00%
Delete error	0.97%	0.90%
Substitute error	1.35%	1.09%
Insert error	0.00%	0.06%

Results are showed in Table 1.

4. ANALYSIS

we can easily see from table 1 that there is only a 0.32% absolute error rate reduction by MGMW HMM method, which is not a big improvement given that the authors claimed a 12% relative error reduction in total digital error rate. In addition, the conditions for the experiments is not clear. First of all, what language is used for digital strings was not state in the paper [1]. If these experiments were done using English, both error rates of the experiments are unacceptable. Secondly, why insert error (should be insertion error), as showed in table 1, is introduced while using MGMW HMM instead of the traditional HMM. What is more, are the training data and evaluation data divided into different catalogs, such as men, women and child? How many states are using for the model and how many mixture of Gaussian are used for each state are not mentioned yet. These factors may affect the result of experiment

dramatically. Less of all above information makes it very hard to evaluate the experiment results and further to convince the effectiveness of this approach.

Another issue for these experiments is the data for initialization. As mentioned by the authors, the initial parameters are essential to the quality of the final continuous density HMM. Also, to obtain the initial parameters of continuous density HMM, some labeled speech data are required. What kind of these labeled data be, and what is their properties are blank. Thus, even with a few positive results, these experiments can't give much support to this new approach.

Furthermore, after we re-examine the algorithms presented earlier carefully, we can find out, even notice straightforwardly, that there are some important theoretical issues are not stated clearly. The most significant thing here is the clustering algorithm. As we known, clustering procedures yield a data description in terms of clusters or groups of data points that possess strong internal similarities. Formal clustering procedures use a criterion function, such as squared distances from the cluster centers, and seek the grouping that maximizes the criterion function. What kind of criterion function is used in MGMW HMM clustering procedure is transparent to the reader, which make the underlying procedure of the this method totally obscure. Another things here is clustering procedure can lead to unmanageable computational problems, how to avoid this kind of problem in MGMW HMM method is not mentioned. Also, intuitively, the number of Gaussian groups are determined by clustering procedure. Why and How the authors get three groups of mixture weights are still a question. Additionally, the relationship between the number of mixture weights vectors and error rates are not analyzed.

On the other hand, although the authors believe that by selecting an appropriate group of mixture weight which outputs maximum probability, we can get a good result for recognition, the correctness of making this choice was not verified or proved.

Overall, by above analysis and with all questions unanswered, the effectiveness of this new method and the value of experiments results are under doubt.

5. SUMMARY

A new method called MGMW HMM was proposed in [1], which claims 12% reduction of error rates in compare with traditional HMM approach with very little increase of computational costs. This paper reviewed the new method. The key equations and ideas of the new method and its experiments results were concisely presented. A detailed analysis of the experiments results and its underlying theory showed that there are still quiet a lot of unclear issues for both of the supporting experiments and the method itself. Insufficient of proof and explanation for this new method make it unacceptable for a practical speech recognition system.

REFERENCES

[1] L. Ming, Y. Tiecheng, "Multi-group Mixture Weight HMM," Proceedings of the 6th International Conference on Spoken Language Processing, vol. 1, pp. 290-292, Beijing, China, October 2000.

[2] R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, Wiley-Interscience, 2001

[3] L. R. Rabiner, B. H. Juang, "Fundamentals of Speech Recognition," Published by PTR Prentice-Hall Inc., 1993.

[4] X. D. Huang, Y. Ariki, M. A. Jack, "Hidden Markov Models For Speech Recognition," Edinburgh University Press, 1990.

[5] Lawrence R. Rabiner. A tutorial on Hidden Markow Models and selected applications in speech recognition. Proceedings of IEEE, 77(2): 257-286, 1989

[6] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," Inequalities, vol. 3, pp. 1-8, 1972.

[7] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," IEEE Trans. on Information Theory, IT-13(2), pp. 260-269, April 1967.