Variational Bayes Adapted GMM Based Models for Audio Clip Classification

Ved Prakash Sahu, Harendra Kumar Mishra, and C. Chandra Sekhar

Speech and Vision Lab, Dept. of Computer Sc. & Engg.,
Indian Institute of Technology-Madras, India
{vedsahu,harendra,chandra}@cse.iitm.ac.in

Abstract. The most commonly used method for parameter estimation in the Gaussian mixture models (GMMs) is maximum likelihood (ML). However, it suffers from the overfitting when the model complexity is high. Adapted GMM is an extended version of GMMs and it helps to reduce the overfitting in the model. Variational Bayesian method helps in determining optimal complexity so that it avoids overfitting. In this paper we propose the variational Bayes learning method for training the adapted GMMs. The proposed approach is free from overfitting and singularity problems that arise in the other approaches. This approach is faster in training and allows a fast-scoring technique during testing to reduce the testing time. Studies on the classification of audio clips show that the proposed approach gives a better performance compared to GMMs, adapted GMMs, variational Bayes GMMs.

Keywords: GMM, variational learning, Bayesian adaptation.

1 Introduction

This paper aims to investigate the behavior of different types of Gaussian mixture based models such as conventional GMMs, adapted GMMs referred to as Gaussian mixture model-universal background model (GMM-UBM), and variational Bayesian GMMs (VB-GMMs) for audio clip classification task. We propose to use the variational Bayes adapted GMMs (VB-GMM-UBM) for audio clip classification and compare it with the GMMs, GMM-UBM, and VB-GMMs.

The GMM commonly uses the maximum likelihood (ML) method for parameter estimation. However, the ML method suffers from the problem of overfitting due to the presence of singularities, when the model complexity is high [1]. One way to handle this is by adapting the parameters of GMM using the training examples. The GMM-UBM [2] is one such technique, where the UBM is built as a large GMM from the data of all classes. The model for a class is obtained by updating the parameters of the UBM using the training examples of the respective class using an adaptation method [2]. This provides a tighter coupling between the class model and the UBM, and gives a better performance than the decoupled models like conventional GMMs. Training a GMM-UBM is much faster than the conventional GMMs and also allows a fast-scoring technique [2] during testing. But it

S. Chaudhury et al. (Eds.): PReMI 2009, LNCS 5909, pp. 513–518, 2009.

suffers from the overfitting when the model complexity is high. To overcome this problem, the VB-GMMs [1] can be used. The VB approach for GMMs helps to determine the optimal number of components to build a model [1]. It prunes the mixture components which are not useful. Hence the variational approach does not suffer from overfitting and has good generalization capabilities [3]. Variational approach assumes that the parameters are not uniquely specified, but instead are modeled by probability density functions. This introduces an additional set of hyperparameters for the distributions of parameters [1].

In this work, we propose to use the VB-GMM-UBM. In the proposed approach, we build the UBM using the variational approach instead of the ML method. Training the UBM using the variational approach gives an optimal number of components in the UBM. The hyperparameters of each component in the UBM are used to derive the Gaussian parameters for the corresponding component in the UBM. The class model is obtained by Bayesian adaptation that adapts the Gaussian parameters of the UBM using the training examples of that class. The proposed approach is observed to be faster in training. It is free from singularity and overfitting problems that arise in the other approaches. This generalizes the model well and hence its performance is better compared to the other models.

Section 2 describes the proposed VB-GMM-UBM approach. Section 3 presents the experimental studies and the summary of the work is presented in section 4.

2 Variational Bayes Adapted GMMs

2.1 Variational Gaussian Mixture Models

The GMMs can be trained using the variational approach. The distribution of d-dimensional feature vectors $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_n, .., \mathbf{x}_N\}$ can be written as a linear superposition of Gaussians in the form $p(\mathbf{x}_n) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\mu_\mathbf{k}, \mathbf{\Lambda}_\mathbf{k}^{-\mathbf{1}})$, where π_k, $\mu_\mathbf{k}$ and $\mathbf{\Lambda_k}$ are the mixture weight, mean vector and precision matrix, respectively for k^{th} Gaussian. In Bayesian approach, we assume priors for the parameters. The conjugate priors are the Dirichlet distribution $Dir(\pi|\alpha)$ for mixture weights, Gaussian distribution $\mathcal{N}(\mu|\mathbf{m}, (\beta\mathbf{\Lambda})^{-\mathbf{1}})$ for mean, and Wishart distribution $\mathcal{W}(\mathbf{\Lambda}|\mathbf{W}, \nu)$ for precision matrix. Here $\{\alpha, \beta, \nu, \mathbf{m}, \mathbf{W}\}$ form a set of hyperparameters. This set of hyperparameters can be obtained using the EM approach [1] as follows:

(E)xpectation Step : Responsibility r_{nk} of the k^{th} component for the n^{th} example $\mathbf{x}_n$ is defined as $r_{nk} = \frac{\rho_{nk}}{\sum_{j=1}^{K} \rho_{nj}}$. Here,

$$\ln \rho_{nk} = d\beta_k^{-1} + \nu_k(\mathbf{x}_n - \mathbf{m}_k)^T \mathbf{W}_k(\mathbf{x}_n - \mathbf{m}_k) + \sum_{i=1}^{d} \psi\left(\frac{\nu_k + 1 - i}{2}\right)$$

$$+ d \ln 2 + \ln |\mathbf{W}_k| + \psi(\alpha_k) + \psi(\sum_{k=1}^{K} \alpha_k), \qquad (1)$$

and ψ is the gamma function.

(M)aximization Step : Hyperparameters are updated as follows:

$$\alpha_k = \alpha_0 + N_k, \qquad \beta_k = \beta 0 + N_k, \qquad \mathbf{m}_k = \frac{1}{\beta 0 + N_k}(\beta_0 \mathbf{m}_0 + N_k \bar{\mathbf{x}}_k),$$

$$\nu_k = \nu_0 + N_k, \qquad \mathbf{W}_k^{-1} = \mathbf{W}_0^{-1} + N_k \mathbf{S}_k + \frac{\beta_0 N_k}{\beta_k}(\bar{\mathbf{x}}_k - \mathbf{m}_0)(\bar{\mathbf{x}}_k - \mathbf{m}_0)^T, \quad (2)$$

where,

$$N_k = \sum_{n=1}^{N} r_{nk}, \qquad \bar{\mathbf{x}}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}\mathbf{x}_n, \qquad \mathbf{S}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}(\mathbf{x}_n - \bar{\mathbf{x}}_k)(\mathbf{x}_n - \bar{\mathbf{x}}_k)^T. \quad (3)$$

Here $N_k, \bar{\mathbf{x}}_k, S_k$ are the intermediate values used to update the hyperparameters. Initial values for priors $\{\alpha_0, \beta_0, \nu_0, \mathbf{m}_0, \mathbf{W}_0\}$ are chosen as suggested in [3].

2.2 Universal Background Model (UBM)

In the VB-GMM-UBM system, the first step is to generate the UBM [2,4]. The UBM is a single and class-independent large GMM. Given N training examples, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, we train the UBM using the VB-EM method described in the previous subsection. This gives the hyperparameters for each mixture component in the UBM. However, Bayesian adaptation needs Gaussian parameters of the mixture components, that are used to build the class specific models by updating the UBM parameters. Hence, the conversion of hyperparameters to the Gaussian parameters is performed in the UBM before the adaptation process starts. In the process of obtaining the Gaussian parameters, we first determine the probabilistic alignment of training examples to the UBM mixture components. The responsibility of a training example $\mathbf{x}_n$ for the k^{th} mixture in the UBM is computed as follows:

$$\zeta_{nk} = \frac{\alpha_k p_k(\mathbf{x}_n)}{\sum_{j=1}^{K} \alpha_j p_j(\mathbf{x}_n)}. \quad (4)$$

Here, $p_k(\mathbf{x}_n)$ is calculated using the student's t-distribution for k^{th} mixture of the UBM as $p_k(\mathbf{x}_n) = \mathrm{St}(\mathbf{x}_n|\mathbf{m}_k, \mathbf{L}_k, \nu_k + 1 - d)$, where $\mathbf{m}_k$ is the mean vector, $\mathbf{L}_k$ is the precision matrix given by $\mathbf{L}_k = (\beta_k(\nu_k + 1 - d)/(1 + \beta_k))\mathbf{W}_k$, and K is the number of mixtures in the UBM. Then we use the responsibility ζ_{nk} to compute the Gaussian parameters such as weight w_k, mean μ_k and variance σ_k^2 of k^{th} component as follows:

$$w_k = \sum_{n=1}^{N} \zeta_{nk}, \qquad \mu_k = \frac{1}{w_k}\sum_{n=1}^{N} \zeta_{nk}\mathbf{x}_n, \qquad \sigma_k^2 = \frac{1}{w_k}\sum_{n=1}^{N} \zeta_{nk}(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^t. \quad (5)$$

2.3 Adaptation of Class Model from UBM

Once the Gaussian parameters of the UBM are computed, the next step is to get the GMM for each class. The class model is obtained by adapting the Gaussian parameters of the UBM using the training examples of the corresponding class [2,5]. Given a UBM and R training vectors $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_R\}$ from a class, we first determine the probabilistic alignment of the training vectors onto the Gaussian mixture components in UBM. For k^{th} mixture in the UBM, the responsibility ξ_{rk} is computed as follows:

$$\xi_{rk} = \frac{w_k p_k(\mathbf{x}_r)}{\sum_{j=1}^{K} w_j p_j(\mathbf{x}_r)}. \tag{6}$$

Then ξ_{rk} and $\mathbf{x}_r$ are used to compute the new parameters as follows:

$$\tilde{w_k} = \sum_{r=1}^{R} \xi_{rk}, \qquad \tilde{\mu_k} = \frac{1}{\tilde{w_k}} \sum_{r=1}^{R} \xi_{rk}\mathbf{x}_r, \qquad \tilde{\sigma_k}^2 = \frac{1}{\tilde{w_k}} \sum_{r=1}^{R} \xi_{rk}(\mathbf{x}_r - \tilde{\mu}_k)(\mathbf{x}_r - \tilde{\mu}_k)^t. \tag{7}$$

The new parameters $\{\tilde{w_k}, \tilde{\mu_k}, \tilde{\sigma_k}^2\}$ obtained from the class specific training data are used to update the old UBM parameters $\{w_k, \mu_k, \sigma_k{}^2\}$ for k^{th} mixture to get the adapted parameters as follows:

$$\hat{w_k} = [\alpha_k^w \tilde{w}_k/R + (1 - \alpha_k^w)w_k]\gamma, \tag{8}$$

$$\hat{\mu_k} = \alpha_k^m \tilde{\mu_k} + (1 - \alpha_k^m)\mu_k, \tag{9}$$

$$\hat{\sigma_k}^2 = \alpha_k^v \tilde{\sigma_k}^2 + (1 - \alpha_k^v)(\sigma_k{}^2 + \mu_k{}^2) - \hat{\mu_k}^2, \tag{10}$$

where $\mu_k{}^2$ and $\hat{\mu_k}^2$ are diagonal elements of matrix $\mu_k \mu_k{}^t$ and $\hat{\mu_k}\hat{\mu_k}^t$ respectively.

The adaptation coefficients $\{\alpha_k^w, \alpha_k^m, \alpha_k^v\}$ control the balance between the old and new parameters for the weights, means, and variances respectively. The scale factor, γ is computed over all adapted mixture weights to ensure that their sum is unity. For each mixture and each parameter, a data-dependent adaptation coefficient $\alpha_k^\rho, \rho = \{w, m, v\}$, is defined as $\alpha_k^\rho = \frac{\tilde{w_k}}{\tilde{w_k} + r^\rho}$, where r^ρ is a fixed relevance factor for parameter ρ. The performance is insensitive to relevance factors in the range $\lfloor 8 - 20 \rfloor$ [2]. We have used a relevance factor of 16. The update of parameters described in Eqs. (8)-(10) can be derived from the general maximum a posteriori (MAP) estimation for a GMM using constraints on the prior distribution described in [6].

In the next section, we study the performance of conventional GMMs, GMM-UBM, VB-GMM and VB-GMM-UBM for audio clip classification task.

3 Studies and Results

In this section we present our studies on audio clip classification. We compare the performance of the proposed approach with that of conventional GMMs, GMM-UBM and VB-GMM.

The audio data set used in our study includes 180 audio clips from 5 categories : advertisement, cartoon, cricket, football, and news from different TV channels, where 120 clips used for training and remaining clips used for testing. The duration of an audio clip is approximately 20 seconds. The 14-dimension features used are clip-based, and are extracted from frame-based features [7]. The features for

Table 1. Audio clip classification accuracy (in %) for the GMMs, GMM-UBM, VB-GMMs, and VB-GMM-UBM

Audio Class	GMMs	GMM-UBM	VB-GMMs	VB-GMM-UBM
Ads	67.21	77.04	68.85	78.68
Cartoon	72.15	91.13	86.07	96.20
Cricket	66.67	70.00	78.33	93.33
Football	57.38	80.32	96.72	96.72
News	73.09	83.33	89.74	91.02
Average	67.30	80.36	83.94	91.18

Table 2. Performance of GMM-UBM and VB-GMM-UBM for adapting different sets of parameters (W=weight, M=mean, V=variance)

Adapted Set of Parameters	Performance(%)	
	GMM-UBM	VB-GMM-UBM
W	15.89	74.36
M	74.85	82.01
V	28.90	73.74
W, M	78.03	90.26
W, V	33.53	86.72
M, V	79.34	89.38
W, M, V	80.36	91.18

clip-based are volume standard deviation, volume dynamic range, volume undulation, 4Hz modulation energy, ZCR standard deviation, nonsilence ratio, pitch standard deviation, similar pitch ratio, nonpitch ratio, frequency centroid, bandwidth, and energy ratio in subbands.

Table 1 shows the classification accuracy of conventional GMMs, GMM-UBM, VB-GMMs, and VB-GMM-UBM on audio clip data. It is seen that the GMM-UBM performs better than the conventional GMMs. In GMM-UBM, adapted class models localize the most unique features for each target audio class and perform better than conventional GMMs. It is also seen that VB-GMM gives a better average performance than the GMM-UBM. This is because the UBM trained by ML method suffers from singularity and overfitting problems. These problems also propagate to the final class model through the adaptation process. Such problems are resolved when the UBM is trained using the variational learning method. The advantage of the variational approach is that the final number of Gaussian components is generally smaller than the initial ones. This is due to the fact that the VB method makes the model selection and parameter learning at the same time. The proposed VB-GMM-UBM model gives 91.18% classification accuracy. The performance is significantly higher compared to 67.30% in GMMs, 80.36% in GMM-UBM and 83.94% in VB-GMMs.

We study the effect of adapting different sets of parameters during generation of the audio class models. Table 2 shows the performance of GMM-UBM and VB-GMM-UBM on adaptation of different sets of Gaussian parameters. It is seen that the mean adaptation plays a more significant role in the GMM-UBM

Table 3. Confusion matrix for VB-GMM-UBM

Audio Class	Classified Audio Class				
	Ads	Cartoon	Cricket	Football	News
Ads	78.68	14.75	1.63	0.0	4.91
Cartoon	2.53	96.20	0.0	0.0	1.26
Cricket	1.66	3.33	93.33	1.66	0.0
Football	1.63	1.63	0.0	96.72	0.0
News	3.84	5.12	0.0	0.0	91.02

than the weight or variance adaptation. In VB-GMM-UBM, each individual parameter adaptation (weight, mean, and variances) plays an important role in adaptation process.

The confusion matrix for the VB-GMM-UBM that gives the best performance is shown in Table 3. It is seen that the Ads category is mainly confused with the Cartoon category and the News category.

4 Summary

This work studies the behavior of different types of Gaussian mixture model based classification models. Conventional GMMs, GMM-UBM, VB-GMMs and the proposed method VB-GMM-UBM are studied. The GMMs and GMM-UBM suffer from the overfitting problem, when the model complexity is high. In such cases, the VB-GMMs perform better. The VB-GMMs prune out the redundant components. The VB-GMMs are also free from singularity problems that arise frequently in GMMs and GMM-UBM. We could improve its performance by applying the adaptation process in VB-GMMs. Results of our studies on audio clip classification indicate that the proposed VB-GMM-UBM model performs better than the other models.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
2. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker verification using adapted Gaussian mixture models. Digital Signal Processing 10, 19–41 (2000)
3. Nasios, N., Bors, A.: Variational learning for Gaussian mixture model. IEEE Trans. System, Man, and Cybernetics 36, 849–862 (2006)
4. Zheng, R., Ulang, S., Xu, B.: Text-independent speaker identification using GMM-UBM and frame level likelihood normalization. In: Proc. ISCSLP, pp. 289–292 (2004)
5. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley, New York (2000)
6. Gauvain, J.L., Lee, C.-H.: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. IEEE Trans. Speech and Audio Processing 2, 291–298 (1994)
7. Aggarwal, G., Bajpai, A., Khan, A.N., Yegnanarayana, B.: Exploring features for audio indexing. Inter-Research Institute Student Seminar, IISc Bangalore (March 2002)