THE INFINITE GAUSSIAN MODELS: AN APPLICATION TO SPEAKER IDENTIFICATION

Souad FRIHA[(1)] & Nora MANSOURI[(2)] & Abdelmalik TALEB AHMED[(3)]

[(1)] Institut de Génie Electrique, University of Tebessa, ALGERIA.
souad_fri@yahoo.com
[(2)] Laboratoire d'Automatique et de Robotique, University of Mentouri, Route Ain EL Bey, Constantine, ALGERIA
nor_mansouri@yahoo.fr
[(3)] University of Valenciennes et du Hainaut Cambrésis LAMIH UMR CNRS-UVHC 8530, Univ. de Valenciennes et du Hainaut Cambrésis, Le mont Houy 59313 Valenciennes Cedex 9 France.
Abdelmalik.Taleb-Ahmed@univ-valenciennes.fr

ABSTRACT

When modeling speech with traditional Gaussian Mixture Models (GMM) a major problem is that one need to fix a priori the number of GMMs. Using the infinite version of GMMs allows to overcome this problem. This is based on considering a Dirichlet process with a Bayesian inference via Gibbs sampling rather than the traditional EM inference. The paper investigates the usefulness of the infinite Gaussian modeling using the state of the art SVM classifiers. We consider the particular case of the speaker identification under limited data condition that is very short speech sequences. Basically, recognition rates of 100% are achieved after only 5 iterations using training and test samples less than 1 second. Experiments are carried out over NIST SRE 2000 corpus.

KEYWORDS: Speaker Identification, Infinite GMM, SVM, Dirichlet Process, Gibbs Sampling

1 INTRODUCTION

We consider the problem of text independent speaker identification. The standard approach to this problem is to model the speaker using a Gaussian Mixture Model (GMM) [2] [4] [6 [10] [12] [17] [26] [34] [35] [36] (34) (35) (36).

From another hand, the Support Vector Machine (SVM) discriminative classifier is currently one among the most robust classifiers in speaker identification and verification [7] [20] as well as in many other areas where data has to be classified in an efficient way [7][15][31][37]. Furthermore, it has been also successfully combined with GMM to increase accuracy [6] [15] [16] [37].

Though dealing with GMMs especially when combined with SVMs is an attractive way to model systems, one has to fix a priori the number of Gaussians involved, then to optimize this number using cross validation technique for example.

This has been an open problem for many years and some research works were carried out in order to estimate the optimal number [1] [22] [25].

Within this work we use an Infinite version of GMMs proposed by Rasmussen in his original paper [28] within a general framework of the so-called Dirichlet Process Mixture (DPM) model. Literature on Dirichlet Process and Gibbs sampling is abundant (see for example [14] [28] [29]).

Despite its important consequence which is overcoming the Gaussian number problem, the Infinite GMM modeling has not been combined to SVM classifiers. And to our knowledge, the IGMM-SVM based method proposed within this paper is a novelty by itself.

For experiments we used very short speech sequences for training and testing which is a practical issue. This is justified by the fact that in many cases in practice there is only limited data for speaker modeling and testing.

According to (Reynolds & Rose, 1995), typically we need utterances longer than 2 s to achieve adequate accuracy in speaker identification [32]. We show that with our proposed hybrid IGMM-SVM method for the speaker identification, this assumption is no longer true. In the present work, limited data denotes the case when having speech data less than one second per utterance.

The outline of the paper is as follows. Sections 2 and 3 describe the basic theoretical framework respectively for SVMs and IGMMs. We present also in section 3 the main principle behind Gibbs sampling. In section 4 we outline jointly the IGMM supervector and the global scheme for our IGMM-SVM speaker identification system. The potential of the approach is verified by applying it to a NIST speaker recognition evaluation 2000 task and comparing it to some results in [32]. This is done by conducting the experiments of Section 5. The high recognition rates obtained show clearly that our new

method is efficient for performing speaker identification in real situations with limited data. Finally, comments are given in a general conclusion.

2 SUPPORT VECTOR MACHINE

An SVM [5] [30] is a two class-classifier based on a hyperplane separators. This separator is chosen in order to maximize the distance between the hyperplane and the closest training vectors. These training vectors are called support vectors. SVMs usually operate in a high dimensional feature space (potentially with infinite dimension), nonlinearly related with a mapping function ϕ .to the original input feature space X. Given an observation $x \in X$ and a mapping function ϕ, an SVM discriminator function is given by :

$$f(x) = \sum_{k=1}^{M} \lambda_k y_k \langle \phi(x), \phi(x_k) \rangle + b \quad (1)$$

$$f(x) = \sum_{k=1}^{M} \lambda_k y_k K(x, x_k) + b \quad (2)$$

The function $K(x_i, x_j)$ is the kernel function and is constrained to have some properties (Mercer conditions) so that $K(.,.)$ can be expressed as :

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \phi(x_i)^T \phi(x_j) \quad (3)$$

Here $\phi(x)$ is a mapping from the input space (where x lives) to a possibly infinite dimensional space. The kernel is required to be positive semi-definite. The Mercer condition ensures that the margin concept is valid and the optimization of the SVM is bounded (5). The "kernel trick" avoids evaluating the mapping function $\phi(.)$ [30].

The x_k 's in [1] are the support vectors and the y_k 's are the corresponding target class values ± 1. M is the number of support vectors and the λ_k 's are obtained through a training process.

Hence, an SVM makes separation with maximal margin in a high dimensional space defined by a kernel function. This is performed by minimizing :

$$\|w\|^2 + C \sum_i \xi_i \quad (4)$$

Subject to:

$$y_i (w\phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (5)$$

C is the penalty parameter which represents the constraint violation of the data points occurring on the wrong side of the SVM boundary ξ_i measures the degree of misclassification of x_i

The three main kernel functions that are often used are the Linear, the Polynomial and the Radial Basis Function (RBF). We used the radial kernel because it's a generalization of the two others and has only 2 adjustable parameters, it is defined as:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (6)$$

(With kernel parameters γ , C)

C is the cost parameter and the parameter γ shows the width of Gaussian functions.

In order to obtain optimal values of these parameters, there is no general rule about the optimization. Some strategies such as grid algorithm or genetic algorithm (8) have been used. This is beyond the scope of this paper and may be a future extension.

3 INFINITE GAUSSIAN MIXTURE MODELING

The infinite GMM belongs to the family of Dirichlet Process Mixtures and can be derived in a number of different ways [33]. A comprehensive discussion of alternative perspectives on the Dirichlet process mixtures can be found in [14] [21] .Within this paper, the concept is introduced through the finite Gaussian mixture model, whose mixing weight is given by a Dirichlet Process prior. The infinite Gaussian mixture model is then derived by considering the situation where the number of mixtures tends to infinity. The inference of the infinite GMM parameters is implemented using Gibbs sampling.

3.1 Finite Gaussian Mixtures

In a Finite Gaussian Mixture Model, the probability density function of data $x = \{x_1, ..., x_n\}$ can be modeled by finite mixtures of Gaussian distributions with k components:

$$p(x / \mu, s, n) = \sum_{j=1}^{k} \pi_j G(\mu_j, s_j^{-1}) \quad (7)$$

Where $\mu = \{\mu_1, ..., \mu_k\}$ are the means, $s = \{s_1, ..., s_k\}$ are the precisions (inverse variances), $\pi = \{\pi_1, ..., \pi_k\}$ are the mixing weights (which must be positive and sum to one) and G is a Gaussian distribution.

The classical approach to estimating the GMM parameters, (μ, s, π) , is to maximize the data likelihood using the expectation-maximization (EM) algorithm. The EM algorithm guarantees to converge to a local maximum, with the quality of the maximum being heavily dependent on the random initialization of the algorithm. On the contrary, the Bayesian approach defies prior distributions over the GMM parameters, and the inference is performed with respect to the posterior probability of the parameters. As opposed to achieving an "optimal" estimate of the parameters, Bayesian inference uses the Monte Carlo method to generate samples from the posterior distribution,

and by averaging over the Monte Carlo samples, the problem of local maximum can be overcome [33].

In general, the priors are specified via "hyper-parameters", which themselves are given higher level priors [33].

The component means are given Gaussian priors:

$$p(\mu_j \mid \lambda, r^{-1}) \tag{8}$$

Where prior mean λ and prior precision r, are hyper-parameters that are common to all components. The hyper-parameters themselves are given vague Gaussian and Gamma hyper-priors :

$$p(\lambda) = G(\mu_x, \sigma_x{}^2) \tag{9}$$

$$p(r) = G_a(1, \sigma_x{}^{-1}) \alpha r^{-\frac{1}{2}} \exp(-r\sigma_x{}^2/2) \tag{10}$$

Where μ_x and σ_x^2 are the mean and the variance of the data points.

To make inferences with respect to component means, the conditional posterior distributions from μ_j are obtained by multiplying the likelihood [7] by the prior [8]), resulting in a Gaussian distribution :

$$p(\mu_j \mid c, x, s_j, \lambda, r) \sim G(\frac{x_j n_j s_j + \lambda r}{n_j s_j + r}, \frac{1}{n_j s_j + r}) \tag{11}$$

Where $\bar{x}_j$ and n_j are the mean and the number of data points belonging to mixture j, respectively. Similarly, the condition posterior for λ and r can be obtained by the multiplication of their likelihoods and hyper-priors to enable Monte Carlo sampling.

The component precisions are given Gamma priors:

$$p(s_j \mid \beta, \omega) \sim Ga(\beta, \omega^{-1}) \tag{12}$$

Where β and ω are again hyper-parameters with priors given by:

$$p(\beta^{-1}) = Ga(1,1) \tag{13}$$

$$p(\omega) = Ga(1, \sigma_x^2) \tag{14}$$

The conditional posterior precisions are obtained by multiplying likelihood and prior:

$$p(s_j \mid c, x, \mu_j, \beta, \omega) \sim$$

$$Ga(\beta + n_j, \frac{\beta + n_j}{\omega\beta + \sum_{i:Ci=j} (x_i - \mu^j)^2})$$

(Here $c = \{c_i, i = 1...n\}$ is introduced to indicate that the data point x_i belongs to mixture c_i. The conditional posteriors for hyper-parameters, β and ω, can also be obtained by multiplying the respective likelihoods and hyper-priors.

As for the general case of mixture models the mixing weights are given Dirichlet priors with concentration parameter $\frac{\alpha}{k}$ [33]

$$p(\pi_1, ..\pi_k / \alpha) \sim dirichlet(\frac{\alpha}{k}, ..., \frac{\alpha}{k}) \tag{16}$$

Sampling for the mixing weights can be indirectly realized by sampling for the indicators, whose probability is conditional on the mixing weights:

$$p(c_1, ...c_k / \pi_1, ...\pi_k) = \prod_{j=1}^{k} \pi_j \tag{17}$$

By integrating out the mixing weights as a result of the properties of the Dirichlet integral, the prior for the indicators is only dependent on α. Furthermore, to use Gibbs sampling for the discrete indicators, c_i, the conditional prior for a single indicator, given all the other indicators, is required and can be obtained as follows:

$$p(c_{i=j} \mid c_{-j}, \alpha) = \frac{n_{-i, j+\alpha/k}}{n - 1 + \alpha} \tag{18}$$

Where the subscript $-i$ indicates all indices except i and $n_{-i,j}$ is the number of data points, excluding x_i, that belong to mixture j. The posteriors are given by the multiplication of the likelihood and the prior:

$$p(c_{i=j} \mid c_{-i}, \mu_j, s_j, \alpha) \propto$$

$$\frac{n_{-i,j} + \frac{\alpha}{k}}{n - 1 + \alpha} s_j{}^{\frac{1}{2}} \exp(-\frac{s_j (x_i - \mu_j)^2}{2}) \tag{19}$$

3.2 Gibbs Sampling

Markov Chain Monte Carlo samples can be generated iteratively to approximate the posteriors for the parameters and hyper-parameters defined above. For finite mixtures of Gaussians, Gibbs sampling proceeds as follows:

1. Sample α ; sample c given new α . Le μ given new λ and r .
2. Sample λ and r ; samp.
3. Sample β and ω ; sample s given new β and ω .
4. Repeat step 1) – 3) until maximum number of iterations is reached.

3.3 Infinite Gaussian Mixtures

The computation with infinite mixtures is finite through the use of "represented" and "unrepresented" mixtures. Represented mixtures are those that have training data associated with them whilst unrepresented mixtures, which are of infinite number, have no training data associated with them. By using unrepresented mixtures, the task of selecting the number of mixtures is avoided.

With the exception of the indicators, the conditional posteriors for the infinite limit, for all the other model parameters/hyper-parameters, are obtained by substituting in k_{rep}, the number of represented mixtures, for k in the above equations. For the indicators, let $k \to \infty$ in (20), and the conditional prior will give the limits:

$$n_{-i,j} \succ 0 : p(c_{i=j} \mid c_{-i}, \alpha) = \frac{n_{-i,j}}{n-1+\alpha} \qquad (20)$$

$$other : p(c_i = other \mid c_{-i}, \alpha) = \frac{\alpha}{n-1+\alpha} \qquad (21)$$

The above priors allow the indicators to be associated with unrepresented mixtures. Therefore there is a finite number of represented mixtures and an infinite number of unrepresented mixtures.

Similarly to the finite mixtures, the posteriors for the indicators are given by:

$$n_{-i,j} \succ 0 : p(c_{i=j} \mid c_{-i}, \mu_j, s_j, \alpha) \propto$$

$$\frac{n_{-i,j}}{n-1+\alpha} s_j^{\frac{1}{2}} \exp\left(-\frac{s_j (x_i - \mu_j)^2}{2}\right)$$

$$other: p(c_i = other \mid c_{-i}, \lambda, r, \beta, \omega, \alpha) = \frac{\alpha}{n-1+\alpha}$$

$$\propto \qquad (22)$$

$$\frac{\alpha}{n-1+\alpha} \int p(x_i \mid \mu_j, s_j) p(\mu_j, s_j \mid \lambda, r, \beta, \omega) d\mu_j ds_j$$

The likelihood with respect to the unrepresented mixtures is an integral over the prior for the mixture parameters. However this integral is not analytically tractable. Rasmussen followed Neal (1998) [28] who suggested to sample from priors (which are Gaussian and Gamma shaped) in order to generate an efficient Monte Carlo sampling strategy to approximate this integral, allowing the number of represented mixtures to vary according to the data along with the MCMC iterations. Therefore the complete sampling procedure for infinite mixtures of Gaussians is similar to that for finite mixtures, except for the sampling for the indicators.

4 MODULAR REPRESENTATION OF THE EXPERIMENTAL PROCESS

Figure 1 illustrates the main steps leading to the construction of the IGMM supervector. In our case the supervector is made of Gaussian's means which model the speaker utterance.
The global scheme of the speaker identification operation is depicted by Figure. 2.

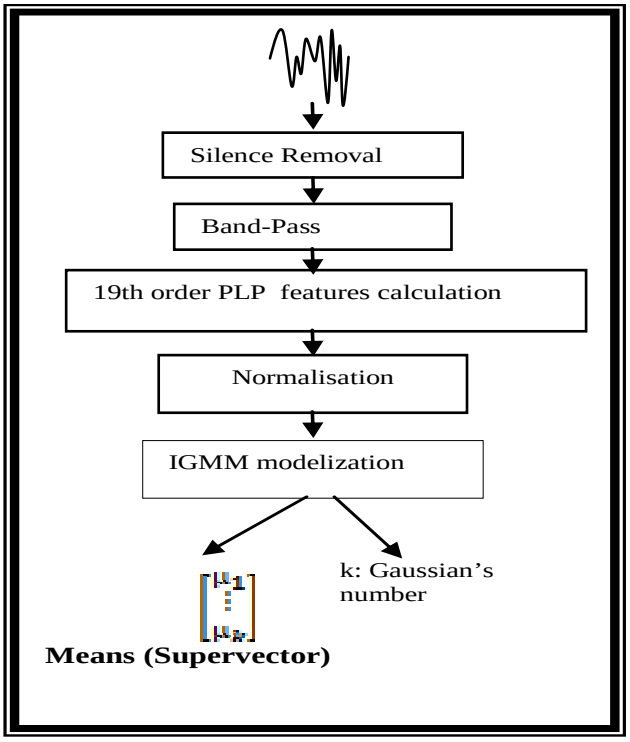

FIgure1: Calculation of GMM supervectors.

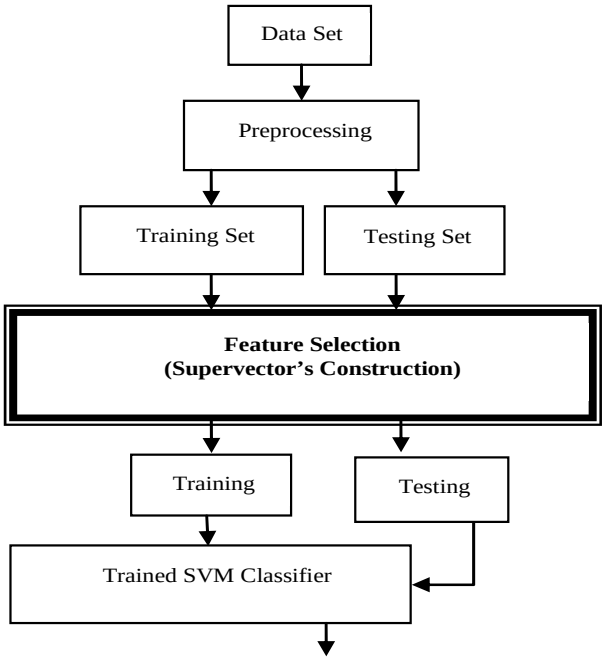

Figure 2: Modular representation of the speaker identification process

5 EXPERIMENTS

We performed experiments on the (2000) NIST data base. TIMIT contains a total of 6300 sentences, 10 sentences spoken by each of 630 speakers from 8 major dialect regions of the United States. All the dialect regions are represented in both subsets, with males and females speakers from each dialect. The amount of overlap of text material in the two subsets is minimized; almost the texts are not identical. All the phonemes are covered in the test material. All phonemes occur multiple times in different contexts.

Roughly 30% of the corpus is used for testing and the remaining 70% for training. In this work, for every speaker 10 recording files are used seven for training and three for testing.

We adopt the assumption in [32] that is in general, one would identify people participating in conversations such as meetings and debates. In our experiments, we assume the maximum number of participants is 20.

We used for the implementation an RBF kernel because it can handle the nonlinearities between class labels and attributes. For the two parameters (C, γ) to be tuned we used the two following values: $C = 10$ and $\gamma = 0.5$ found in the literature [29] as yielding to good results for speech applications.

In the following we describe and comment the different experiences performed to underline the advantages of the new IGMM-SVM method. First, we return back to the over-fitting problem then we present the results of experiments showing the evolution of the recognition rate according to the amount of data and the number of iterations respectively.

5.1 Recognition rate according to the amount of data

The GMM needs sufficient data to model the speaker well to yield good recognition rate [30] . So we were interested in investigating the effectiveness of our IGMM-SVM based method for very short speech sequences.

The significance of the amount of speech data for speaker modeling and testing has been studied for example by Prasanna & al (2006) [30] and it has been demonstrated experimentally that when the speech data for training is less, then the performance is poor due to poor speaker modeling and insufficient speech data to make reliable decision during testing. This is illustrated by Figure.3 where is depicted the recognition rate according to the amount of training/testing data. The results are obtained by using progressive amounts of data for each speaker. For this purpose three training/testing sets are used. The first set corresponds to five females within the first dialect (New England region). The second set corresponds to ten speakers from the first dialect (five females and five males) and the third one concerns ten male speakers equally distributed between the first and the second dialect (Northern region).

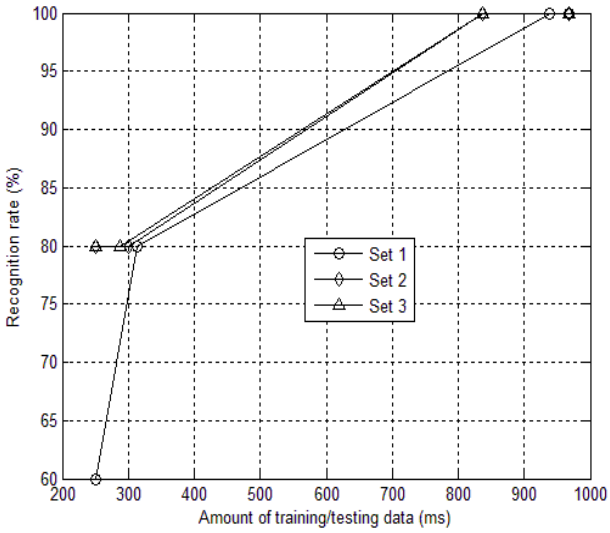

Figure. 3: Speaker recognition rates for different sizes of training and testing data less than 1 second

We notice the recognition rate is better for the set 2 and 3 than for set 1 corresponding to 250ms. This is due to the fact that the whole data belonging to ten speakers is much more than the one belonging to only five speakers and hence the recognition rate must be greater.

A comparable work in [32] investigating the relation between the error rate and the length of short utterances, was conducted over utterances less than 2s. Three speaker-identification methods (conventional GMM, GMM with LDA, and a new method) were applied on various lengths of speech data (0.25, 0.5, 1, and 2-s spontaneous utterances). The best recognition rate in [32] was 85% accuracy while with our method 100% recognition rate is achieved with less than 1s-term utterance.

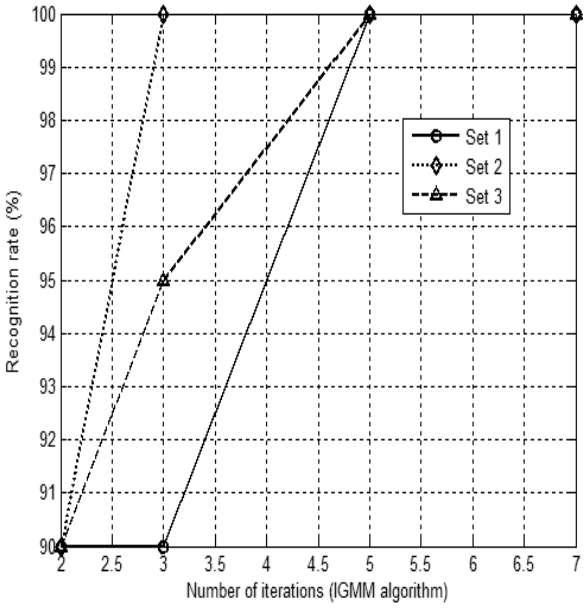

Figure. 4: Evolution of the recognition rate according to IGMM iterations (20 speakers)

Apart this experience, in all the remaining of the paper, the results are obtained after 5 iterations.

5.2 Recognition rate according to the number of iterations

We have noticed that convergence is reached after a small number of iterations namely 5 iterations. Figure 4 illustrates the recognition rate according to the iterations number of the IGMM algorithm for three training/testing sets. The first set corresponds to 10 speakers taken from the first dialect, the second set corresponds to 10 speakers taken from the third dialect (North Midland region) and the third one corresponds to 20 female speakers equally chosen from four dialects (the fourth dialect belongs to the South Midland region).

6 CONCLUSION

This paper explored the effectiveness of combining the Infinite GMM to SVM in a speaker recognition context.

Recognition rates of 100% were obtained after only 5 iterations using training and test samples less than 1second.

Gaussians is used for the modeling. This is not the case with the traditional GMM modeling based on maximum likelihood where often the cluster number or the dimensionality of the feature vector is increased in order to increase the likelihood, leading to an over tuning of the parameters.

The method is hence useful for detecting speakers from very short segments in speech indexing applications as well as for improved performance for rapid speaker identification. The method promises also good performance for longer data segments. This will be a future extension of the present work. Another interesting perspective will be the investigation into the tuning of the SVM parameters in order to optimize the convergence as well as the recognition rates.

The good results obtained are mainly due to the fact that with Infinite GMM, only the right number of Gaussians is used for the modeling. This is not the case with the traditional GMM modeling based on maximum likelihood where often the cluster number or the dimensionality of the feature vector is increased in order to increase the likelihood, leading to an over tuning of the parameters.

The method is hence useful for detecting speakers from very short segments in speech indexing applications as well as for improved performance for rapid speaker identification. The method promises also good performance for longer data segments. This will be a future extension of the present work. Another interesting perspective will be the investigation into the tuning of the SVM parameters in order to optimize the convergence as well as the recognition rates.

BIBLIOGRAPHIE

[1] M. F. Abu El yazid, M. A. El Gamal,, M. M. H. El Ayadi, “On the Determination of Optimal Model Order for GMM-Based Text-Independent Speaker Identification”, EURASIP Journal on Applied Signal Processing, 2004.

[2] Baibo Zhang, Changshui Zhang, Xing Yi, “Active Curve Axis Gaussian Mixture Models”, Pattern Recognition Society Journal 2351-2362, 38, 2005.

[3] Christophe Biernacki, “Pourquoi les modèles de mélange pour la classification?”,Revue MODULAD, No. 40, 2009.

[4] Bin Ma, Donglai Zhu, Rong Tong and Haizhou Li, “Speaker Cluster Based GMM Tokenization for Speaker Recognition”, Interspeech, Pensylvania, USA pp. 505-508, September 2006.

[5] W. M. Campbell, D. E. Sturim, D.A. Reynolds, “Support Vector Machines using GMM Supervectors for Speaker Verification”, IEEE Signal Processing Letters, pp. 308-311, vol. 13, no. 5, 2006.

[6] M. Campbell, D. E. Sturim, D.A. Reynolds, A. Solomonoff, “SVM Based Speaker Verification using a GMM Supervector Kernel and NAP Variability Compensation”, pp. 97-100, in Proc. ICASSP. Toulouse : IEEE, 2006.

[7] Chang Huai You, Kong Aik Lee and Haizhou Li, “An SVM Kernel with GMM-Supervector Based in the Bhattacharryya Distance for Speaker Recognition”, IEEE Signal Processing Letters, Vol. 16, No. 1 January 2009.

[8] Cheng-Lung Huang a, Jian-Fan Dun, “A distributed PSO–SVM hybrid system with feature selection and parameter optimization”, Applied Soft Computing, 1381–1391, 8, 2008

[9] Cheng-Lung Huang, Chieh-Jen Wang, “A GA-based feature Selection and Parameters Optimization for Support Vector Machines”, Expert Systems with Applications, 231-240, 31, 2006.

[10] Dalei Wu, Ji Li, Haiqing Wu, “”-Gaussian Mixture Modelling for Speaker Recognition”, Pattern Recognition Letters 589-594, 30, 2009.

[11] Dalei Wu, Ji Li and Hui Jiang “Normalization and Transformation Techniques for Robust Speaker Recognition”, Dalei Wu, Baojie Li and Hui Jiang, Speech Recognition, Technologies and Applications, pp. 550, ISBN 978-953-7619-29-9, November 2008.

[12] Douglas A. Reynolds, Thomas F. Quatieri and Robert B. Dunn, “Speaker Verification Using Adapted Gaussian Mixture Models”, Digital Signal Processing, pp. 19-41, vol. 10, no. 1-3, 2000

[13] Edmund Jackson, Manuel Davy, Arnaud Doucert, William J. Fitzgerald, “Bayesian Unsupervised Signal Classification by Dirichlet Process Mixtures of Gaussian Processes”, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007.

[14] Edmund Jackson, Manuel Davy and William J. Fitzgerald, "Unsupervised Classification of Functions using Dirichlet Process Mixtures of Gaussian Processes", CUED/F-INFENG/TR. 562, august, 2006.

[15] Emad Bahrami, Samani, M. Mehdi Homayounpour and Hong Gu, "A Novel Hybrid GMM/SVM Architecture for Protein Secondary Structure Prediction", 7th international workshop on Fuzzy Logic and Applications: Applications of Fuzzy Sets Theory, pp. 491-496, LNAI 4578, 2007.

[16] Frank Wood, Michael J. Black, "A nonparametric Bayesian alternative to spike sorting", Journal of Neuroscience Methods 173, 1-12, 2008.

[17] Friha, S.; Mansouri, "Application of GMMs to Speech Recognition using very short time series", Intelligent Systems and Automation: 1st Mediterranean Conference on Intelligent Systems and Automation (CISA 08). AIP Conference Proceedings, pp. 450-453, Volume 1019, 2008.

[18]] Ganapathiraju A., Hamaker J., Picone J., "Applications of Support Vector Machines to Speech Recognition", IEEE Transactions on Signal Processing, vol. 52, no. 8, pp. 2348-2355, 2004.

[19] H S Jayanna and S R Mahadeva Prasanna, "An Experimental Comparison of Modelling Techniques for Speaker Recognition under Limited Data Condition", Sadhana, Indian Academy of Sciences, pp. 717-728, Vol. 34, Part 3, October 2009.

[20] Jérôme Louradour, Khalid Daoudi and Francis Bach, "Feature Space Mahalanobis Sequence Kernels: Application to SVM Speaker Verification", IEEE Transactions on Audio Speech and Langage Processing, Vol. 15, No. 8, November 2007.

[21]] Jian Wang, Jianjun Lei, Jun Guo and Zhen Yang, "SVM Based Speaker Selection using GMM Supervector for Rapid Speaker Adaptation", 6th international conference, SEAL proceedings (Hefei, China), pp. 617-624, vol. 4247, 2006.

[22] Mckenma S. J., Nait Charif H. "Summerizing Contextual Activity and Detecting Unusual Inactivity in a Supportive ", Home Environnement, Pattern Annal Appl. 7(4), 386-401, 2004.

[23] Michael Mendel, "Implementing the Infinite GMM", within the project in Tony Jebara's Machine Learning course, cs4771,may 6, 2005, downloadable from http://mr-pc.org/work/

[24] David M. Blei, Michael I. Jordan, Thomas L. Griffiths, Joshua B.Tenenbaum, "Hierarchical Topic Models and the Nested Chinese Restaurant Process", Advances in Neural Information Processing Systems (NIPS) 16, 2004.

[25] Oh-Wook Kwon, Kwokleung Chan, Te-Won Lee, "Speech Feature Analysis Using Variational Bayesian PCA", IEEE Signal Processing Letters, Vol. XX, No. Y, Month 2002.

[26] Pablo Martine-Olmos, Juan José Murillo-Fuentes and Fernando Pérez-Cruz, "Soft LDPC Decoding in nonlinear Channels with Gaussian Process for Classification", 17th European Signal Processing Conference (EUSIPCO 2009).

[27] Radford M. Neal, "Markov Chain Sampling Methods for Dirichlet Process Mixture Models", Technical Report No.9815, Department of Statistics, University of Torento, 1998.

[28] Carl Edward Rasmussen, in Advances in Neural Information Processing Systems 12, pp. 554-560, MIT Press, 2000.

[29] Carl Edward Rasmussen and Zoubin Ghahramani, "Infinite Mixtures of Gaussian Process Experts", In Advances in Neural Information Processing Systems 14, MIT Press, 2002.

[30] Réda Dehak, Najim Dehak, Patrick Kenny, Pierre Dumouchel, "linear and nonlinear Kernel GMM SuperVector Machines for Speaker Verification", in Interspeech, Antwerp, Belgium, 2007.

[31] Shui-Sheng Zhou, Hong-Wei Liu, Feng Ye, "Variant of Gaussian Kernel and Parameter setting Method for Nonlinear SVM ", Elsevier Science Publishers B. V, Neurocomputing, Issue 13-15, Volume 72, August 2009.

[32] Soonil Kwon a, Shrikanth Narayanan, "Robust speaker identification based on selective use of feature vectors", Pattern Recognition Letters, 85–89, 28, 2007.

[33] Tao Chen, Julian Morris and Elaine Martin, "Bayesian Control Limits for Statistical Process Monitoring", IEEE International Conference on Control and Automation (Budapest, Hungary), ICCA 2005.

[34] Tingyao Wu, Jacques Duchateau, Jean Pierre Martens, Dirk Van Compernolle, "Feature Subset Selection for Improved Native Accent Identification", Speech Communication, 83-98, 52, 2010.

[35] Tomi Kinnunen, Haizhou Li, "An Overview of Text-Independent Speaker Recognition: from Features to Supervectors", Speech Communication, 12–40,52, 2010.

[36] William J.J. Roberts and Jonathan P. Willmore, "Automatic Speaker Recognition using Gaussian Mixture Models", Commonwealth of Australia, 1999.

[37] Xiaodan Zhuang, Jing Huang, Gerasimos Potamiamos, Mark Hasegawa-Johnson, "Acoustic Fall Detection using Gaussian Mixture Models and GMM Supervectors", IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.