An SVM Kernel With GMM-Supervector Based on the Bhattacharyya Distance for Speaker Recognition

Chang Huai You, *Member, IEEE*, Kong Aik Lee, *Member, IEEE*, and Haizhou Li, *Senior Member, IEEE*

***Abstract*—Gaussian mixture model (GMM) and support vector machine (SVM) have become popular classifiers in text-independent speaker recognition. A GMM-supervector characterizes a speaker's voice with the parameters of GMM, which include mean vectors, covariance matrices, and mixture weights. GMM-supervector SVM benefits from both GMM and SVM frameworks to achieve the state-of-the-art performance. Conventional Kullback–Leibler (KL) kernel in GMM-supervector SVM classifier limits the adaptation of GMM to mean value and leaves covariance unchanged. In this letter, we introduce the GMM-UBM mean interval (GUMI) concept based on the Bhattacharyya distance. This leads to a new kernel for SVM classifier. Comparing with the KL kernel, the new kernel allows us to exploit the information not only from the mean but also from the covariance. We demonstrate the effectiveness of the new kernel on the 2006 National Institute of Standards and Technology (NIST) speaker recognition evaluation (SRE) dataset.**

***Index Terms*—Gaussian mixture model, National Institute of Standards and Technology (NIST) evaluation, speaker recognition, supervector, support vector machine.**

I. Introduction

SPEAKER recognition is the process of validating a claimed identity by evaluating the extent to which a test utterance matches the claimant's model. In text-independent speaker recognition, both Gaussian mixture model (GMM) [1] and support vector machine [2] have been proven to be effective and most popularly used for many years.

In a GMM system, universal background model (UBM) is usually trained through expectation-maximization (EM) algorithm by using background data that include a wide range of speakers, languages (for multilanguage application), communication channels, recording devices, and environments. GMM-UBM becomes a popular recognizer in the field of text-independent speaker recognition for its reliable performance reported in the literature. In the GMM approach, the speaker model is obtained by adapting the speaker GMM from UBM through maximum *a posteriori* (MAP) criterion.

Because of the effectiveness of the GMM in modeling the characteristics of speaker, recently, the GMM-supervector that consists of the normalized mean vector of the Gaussian mixture components is adopted to characterize speech utterances. This allows the discriminative classification to be applied in the high-dimensional space. As a result, the kernel of SVM being of an inner product of the high-dimensional vectors can be appropriately exploited for the similarity estimation. The critical factor of SVM in speaker recognition is to design a proper kernel metric. In [3], Campbell *et al.* constructed an SVM kernel using the GMM-supervector that is formed by using the parameters of GMM. They combine the SVM with the GMM-supervector concept and derive an SVM kernel based on Kullback–Leibler (KL) divergence [4] between two GMMs.

The Bhattacharyya distance has found widespread use in numerous applications and has shown to give better results than the KL divergence [5]. In this letter, we introduce a GMM-UBM mean interval (GUMI) vector concept based on the Bhattacharyya distance. Consequently, we propose a GMM-supervector SVM kernel with GUMI. GMM-supervector creates a bridge to transfer the information of the utterance to the parameters of GMM, which is composed of mean, covariance, and mixture weight. One of the transferring methods is the MAP adaptation from UBM. Comparing with the KL kernel in [3], the new kernel provides the potential possibility to exploit the information not only from the mean but also from the covariance. The validity to exploit covariance information is recently proven in a language recognition system reported in [6]. We compare the proposed kernel with the linear KL kernel under exactly the same experimental conditions and follow the testing task specified by the NIST SRE [7]. In the performance evaluation, we also apply strictly the same implementation procedure of nuisance attribute projection (NAP) for both conventional and proposed kernels. In the remainder of this letter, we review the KL kernel in Section II. We derive the new kernel based on the Bhattacharyya distance in Section III. The performance evaluation is shown in Section IV. Finally, we summarize this letter in Section V.

II. Conventional GMM-SVM

A. GMM-Supervector

The UBM is trained using the background databases that are selected to reflect the alternative imposter speeches. The EM algorithm is used for the UBM training. The GMM probability density can be described as follows:

$$p(\mathbf{x}) = \sum_{i=1}^{M} \omega_i f(\mathbf{x}\,|\,\mathbf{m}_i, \Sigma_i) \quad (1)$$

where $\mathbf{x}$ is a D-dimensional cepstral feature vector, and $\mathbf{m}_i, \Sigma_i, \omega_i, (i = 1, \ldots, M)$ are, respectively, the mean vector, the covariance matrix, and the weight of the ith Gaussian component. $f(\cdot)$ denotes Gaussian density function, i.e.,

$$f(\mathbf{x}\,|\,\mathbf{m}_i, \Sigma_i) = \frac{(2\pi)^{-D/2}}{|\Sigma_i|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T \Sigma_i^{-1}(\mathbf{x}-\mathbf{m}_i)\right). \quad (2)$$

Manuscript received June 04, 2008; revised August 08, 2008. Current version published December 30, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Steve Renals.

The authors are with the Institute for Infocomm Research (I2R) Agency for Science, Technology and Research (A*STAR), Singapore 138632 (e-mail: echyou@i2r.a-star.edu.sg; kalee@i2r.a-star.edu.sg; hli@i2r.a-star.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2008.2006711

1070-9908/$25.00 © 2008 IEEE

Authorized licensed use limited to: National University of Singapore. Downloaded on January 7, 2009 at 01:15 from IEEE Xplore. Restrictions apply.

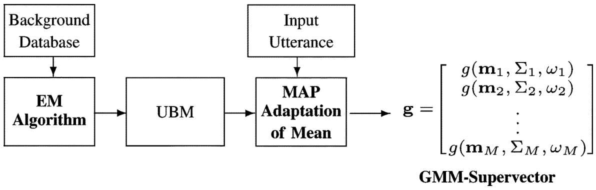

Fig. 1. Process of generating the GMM-supervector from an utterance. $g(\mathbf{m}_i, \Sigma_i, \omega_i)$ is the function that represents the normalized mean aligned by covariance and weight.

The UBM can be expressed by

$$\mathrm{u} = \left\{ \omega_i^{(\mathrm{u})}, \mathbf{m}_i^{(\mathrm{u})}, \Sigma_i^{(\mathrm{u})} \middle| i = 1, 2, \ldots, M \right\}. \quad (3)$$

The speaker GMM, λ, can be obtained by MAP adaptation, and it has the same form as follows:

$$\lambda = \left\{ \omega_i^{(\lambda)}, \mathbf{m}_i^{(\lambda)}, \Sigma_i^{(\lambda)} \middle| i = 1, 2, \ldots, M \right\}. \quad (4)$$

The process of generating the GMM-supervector can be summarized in Fig. 1. The GMM-supervector is formed by concatenating the normalized means of the Gaussian components [8].

B. KL Divergence Kernel

GMM-supervector SVM combines both generative and discriminative methods and leads to the generative SVM kernels based on the probability distribution estimation. Recently, GMM-supervector SVM with KL kernel [3] achieves state-of-the-art performance in speaker recognition. In conventional GMM-supervector SVM system, KL divergence is used to measure the distance of the two GMMs. The KL divergence between the two distributions p_a and p_b is given by

$$\Psi_{\mathrm{KL}}(p_a \,\|\, p_b) = \int_{R^n} p_a(\mathbf{x}) \log\left(\frac{p_a(\mathbf{x})}{p_b(\mathbf{x})}\right) d\mathbf{x}. \quad (5)$$

Let p_a and p_b denote the two probability distributions of GMM_a and GMM_b. By bounding the divergence with the log-sum inequality [8] and assuming that the mixture weight can be ignored, the divergence is constrained by

$$\begin{aligned}\Psi_{\mathrm{KL}}(p_a \,\|\, p_b) &\leq \sum_{i=1}^{M} \Psi_{\mathrm{KL}}(p_{a_i} \,\|\, p_{b_i}) \\ &= \sum_{i=1}^{M} \omega_i \Psi_{\mathrm{KL}}\left(f\left(\mathbf{m}_i^{(\mathrm{a})}, \Sigma_i^{(\mathrm{a})}\right) \middle\| \right. \\ &\qquad \left. \times f\left(\mathbf{m}_i^{(\mathrm{b})}, \Sigma_i^{(\mathrm{b})}\right)\right). \quad (6)\end{aligned}$$

With the assumption that the covariance adaptation can also be ignored, the KL linear kernel function is obtained as follows:

$$K_{\mathrm{KL}}(X_a, X_b) = \sum_{i=1}^{M} \left(\sqrt{\omega_i} \Sigma_i^{-\frac{1}{2}} \mathbf{m}_i^{(\mathrm{a})}\right)^T \left(\sqrt{\omega_i} \Sigma_i^{-\frac{1}{2}} \mathbf{m}_i^{(\mathrm{b})}\right) \quad (7)$$

where X_a and X_b denote the respective feature vector sequences of utterances a and b.

III. Proposed GMM-SVM Kernel

Since the KL divergence is not inner-product symmetric, in the derivation of a kernel, it hinges on the assumption that the covariance is not adapted [3], [8]. In various signal selection problem, the Bhattacharyya distance has shown to outperform the divergence. In this letter, we adopt the Bhattacharyya distance to measure the degree of similarity between two probability distributions. We propose the Bhattacharyya mean distance to represent the similarity of two GMMs. Subsequently, a new kernel is derived which involves the adaptation of mean and covariance.

The Bhattacharyya distance [5] between two probability distributions p_a and p_b is defined by

$$\Psi_{\mathrm{Bhatt}}(p_a \,\|\, p_b) = -\ln\left(\int_{R^n} \sqrt{p_a(\mathbf{x})}\sqrt{p_b(\mathbf{x})}\, d\mathbf{x}\right). \quad (8)$$

So the Bhattacharyya distance for the two GMMs are given by

$$\Psi_{\mathrm{Bhatt}}(p_a \,\|\, p_b) = -\ln\left(\int_{R^n} \sqrt{\sum_{i=1}^{M} p_{a_i}(\mathbf{x})}\sqrt{\sum_{i=1}^{M} p_{b_i}(\mathbf{x})}\, d\mathbf{x}\right) \quad (9)$$

and the Bhattacharyya distance between the ith component of GMM_a and the one of GMM_b is obtained

$$\begin{aligned}\Psi_{\mathrm{Bhatt}}(p_{a_i} \,\|\, p_{b_i}) &= \frac{1}{8}\left(\mathbf{m}_i^{(\mathrm{b})} - \mathbf{m}_i^{(\mathrm{a})}\right)^T \left[\frac{\Sigma_i^{(\mathrm{a})} + \Sigma_i^{(\mathrm{b})}}{2}\right]^{-1} \\ &\quad \times \left(\mathbf{m}_i^{(\mathrm{b})} - \mathbf{m}_i^{(\mathrm{a})}\right) \\ &\quad + \frac{1}{2}\ln\frac{\left|\frac{\Sigma_i^{(\mathrm{a})} + \Sigma_i^{(\mathrm{b})}}{2}\right|}{\sqrt{\left|\Sigma_i^{(\mathrm{a})}\right|\left|\Sigma_i^{(\mathrm{b})}\right|}} - \frac{1}{2} ln\left(\omega_i^{(\mathrm{a})}\omega_i^{(\mathrm{b})}\right). \quad (10)\end{aligned}$$

Bounding (9) with the log-sum inequality, we have

$$\begin{aligned}\Psi_{\mathrm{Bhatt}}(p_a \| p_b) &\leq \sum_{i=1}^{M} \Psi_{\mathrm{Bhatt}}(p_{a_i} \| p_{b_i}) \\ &= -\sum_{i=1}^{M} \ln\left[\int_{R^n} \sqrt{\omega_i^{(\mathrm{a})} f\left(\mathbf{x} \middle| \mathbf{m}_i^{(\mathrm{a})}, \Sigma_i^{(\mathrm{a})}\right)}\right. \\ &\quad \left. \times \sqrt{\omega_i^{(\mathrm{b})} f\left(\mathbf{x} \middle| \mathbf{m}_i^{(\mathrm{b})}, \Sigma_i^{(\mathrm{b})}\right)}\, d\mathbf{x}\right] \\ &= \frac{1}{8}\sum_{i=1}^{M} \left\{\left(\mathbf{m}_i^{(\mathrm{b})} - \mathbf{m}_i^{(\mathrm{a})}\right)^T \left[\frac{\Sigma_i^{(\mathrm{a})} + \Sigma_i^{(\mathrm{b})}}{2}\right]^{-1}\right. \\ &\quad \left. \times \left(\mathbf{m}_i^{(\mathrm{b})} - \mathbf{m}_i^{(\mathrm{a})}\right)\right\} \\ &\quad + \frac{1}{2}\sum_{i=1}^{M}\left[\ln\frac{\left|\frac{\Sigma_i^{(\mathrm{a})} + \Sigma_i^{(\mathrm{b})}}{2}\right|}{\sqrt{\left|\Sigma_i^{(\mathrm{a})}\right|\left|\Sigma_i^{(\mathrm{b})}\right|}}\right] \\ &\quad - \frac{1}{2}\sum_{i=1}^{M} ln\left(\omega_i^{(\mathrm{a})}\omega_i^{(\mathrm{b})}\right). \quad (11)\end{aligned}$$

The right-hand side of (11) consists of three terms. The first term reflects the degree of mean statistical similarity; the second

Authorized licensed use limited to: National University of Singapore. Downloaded on January 7, 2009 at 01:15 from IEEE Xplore. Restrictions apply.

represents the degree of consistency of the covariance matrices; and the third term is the weighting factor. The first term is the most informative quantity in measuring the similarity between the two GMMs[1]. The mean statistics gives the major characteristics of the adaptation; we thus adopt the Bhattacharyya mean distance to represent the similarity of the two GMMs, i.e.,

$$\Phi(p_a \,\|\, p_b) = \sum_{i=1}^{M} \left\{ \left(\mathbf{m}_i^{(\mathrm{b})} - \mathbf{m}_i^{(\mathrm{a})}\right)^T \times \left[\frac{\Sigma_i^{(\mathrm{a})} + \Sigma_i^{(\mathrm{b})}}{2}\right]^{-1} \left(\mathbf{m}_i^{(\mathrm{b})} - \mathbf{m}_i^{(\mathrm{a})}\right) \right\}. \quad (12)$$

Motivated by (12), we define the GMM-UBM mean interval (GUMI) vector to represent the interval between the mixture of the GMM and the corresponding one of UBM as follows:

$$\mathbf{s}_i^{(\lambda)} = \left[\frac{\Sigma_i^{(\lambda)} + \Sigma_i^{(u)}}{2}\right]^{-\frac{1}{2}} \left(\mathbf{m}_i^{(\lambda)} - \mathbf{m}_i^{(\mathrm{u})}\right) \qquad i = 1, \ldots, M \quad (13)$$

where $\mathbf{m}_i^{(\lambda)}$ and $\mathbf{m}_i^{(\mathrm{u})}$ are the mean vectors of the ith mixture of the GMM and UBM, respectively. The GUMI supervector is the concatenation of the GUMI vectors, i.e., $\mathbf{S} = [\mathbf{s}_1^T\, \mathbf{s}_2^T \cdots \mathbf{s}_M^T]^T$. Subsequently, the Bhattacharyya mean distance of a GMM and UBM is obtained by

$$\begin{aligned}\Phi\left(p^{(\lambda)} \middle\| p^{(\mathrm{u})}\right) &= \sum_{i=1}^{M} \left(\mathbf{s}_i^{(\lambda)}\right)^T \mathbf{s}_i^{(\lambda)} \\ &= \left(\mathbf{S}^{(\lambda)}\right)^T \mathbf{S}^{(\lambda)}. \end{aligned} \quad (14)$$

Equation (14) can also be understood as the Bhattacharyya mean distance of a GUMI supervector $\mathbf{S}^{(\lambda)}$. Obviously, the Bhattacharyya mean distance of GUMI supervector $(\mathbf{S}^{(a)} - \mathbf{S}^{(b)})$ is given by

$$\Phi\left(p^{(a-b)} \middle\| p^{(\mathrm{u})}\right) = \left(\mathbf{S}^{(a)} - \mathbf{S}^{(b)}\right)^T \left(\mathbf{S}^{(a)} - \mathbf{S}^{(b)}\right). \quad (15)$$

Therefore, we propose a kernel to be a linear combination of the Bhattacharyya mean distances

$$\begin{aligned} &K_{\mathrm{GUMI}}(\mathbf{X}_a, \mathbf{X}_b) \\ &= \frac{1}{2}\left[\Phi\left(p^{(a)} \middle\| p^{(\mathrm{u})}\right) + \Phi\left(p^{(b)} \middle\| p^{(\mathrm{u})}\right) - \Phi\left(p^{(a-b)} \middle\| p^{(\mathrm{u})}\right)\right] \\ &= \left(\mathbf{S}^{(\mathrm{a})}\right)^T \mathbf{S}^{(\mathrm{b})} \\ &= \sum_{i=1}^{M} \left\{ \left[\left(\frac{\bar{\Sigma}_i^{(\mathrm{a})} + \bar{\Sigma}_i^{(\mathrm{u})}}{2}\right)^{-\frac{1}{2}} \left(\mathbf{m}_i^{(\mathrm{a})} - \mathbf{m}_i^{(\mathrm{u})}\right)\right]^T \times \left[\left(\frac{\bar{\Sigma}_i^{(\mathrm{b})} + \bar{\Sigma}_i^{(\mathrm{u})}}{2}\right)^{-\frac{1}{2}} \left(\mathbf{m}_i^{(\mathrm{b})} - \mathbf{m}_i^{(\mathrm{u})}\right)\right] \right\}. \end{aligned} \quad (16)$$

[1]In conventional GMM-supervector strategy, it is assumed that the covariance and mixture weight are not changed. From the assumption of non-adaptation of the covariance and weight in conventional KL kernel, we may arrive at a conclusion that the information of covariance and weight is not so important as the mean. In other words, it means that the second and third terms are not so important as the first term in the right part of (11).

The kernel is an inner product of the GUMI supervector. It reflects the inter-speaker distance of two GMM distributions by taking out the effect of speaker-independent property represented by UBM. Obviously, it also satisfies the Mercer condition [9].

Since each element of the GUMI supervector is Gaussian distributed, with the Bayesian minimum risk criterion, the kernel scoring can be obtained by

$$\begin{aligned}\Gamma_{\mathrm{cost}}(\mathbf{X}) &= \sum_{l=1}^{L} \alpha_l t_l K_{\mathrm{GUMI}}(\mathbf{X}_l, \mathbf{X}) + d \\ &= \left(\sum_{l=1}^{L} \alpha_l t_l \mathbf{S}^{(\mathbf{X}_l)}\right)^T \mathbf{S}^{(\mathbf{X})} + d \\ &= \mathbf{w}^T \mathbf{S}^{(\mathbf{X})} + d \end{aligned} \quad (17)$$

where t_l is the target value of $+1$ or -1 corresponding to the target class or non-target class, in GMM-supervector, $\mathbf{X}_l$ is actually a sequence of feature vectors of utterance l. $\alpha_l > 0$ is the weight of the vector $\mathbf{X}_l$ so that $\sum_{l=1}^{L} \alpha_l t_l = 0$. d is a bias parameter independent of the observed sequence.

Given a set of linearly separable two-class data, there are many possible solutions. The SVM is a binary linear classifier represented by a hyperplane separator. The separator is selected by maximizing the distance between the hyperplane and the closest training vectors. By introducing SVM, the $\mathbf{X}_l|_{l=1,\ldots,L}$ are selected from the training vector sequences and called the support vectors since they support the hyperplanes on both sides of the margin, and $\mathbf{w}$ is the linear combination of the support vectors. The support vectors are obtained by a quadratic optimization [10]. With the trained model represented by parameters $\mathbf{w}$ and d, the cost value in (17) is used as the score during recognition.

In [6], Campbell extends the KL kernel by using the symmetrized version of the KL divergence. The extended KL kernel contains two terms, i.e., mean vector term and covariance term. Actually, the mean vector term is the same as the conventional KL one. In other words, the adaptation of covariance cannot be used for the mean vector term in the extended KL kernel. This is different from our proposed GUMI kernel. The covariance term relates to the information of covariance, where the adaptation of covariance can be used. However, in the extended KL kernel, the introduction of the covariance term doubles the dimension of the GMM-supervector.

IV. Performance Evaluation

The performance evaluation is conducted on the 2006 NIST SRE core test [7], [12], where 51 448 trials are tested, which includes 3612 true trials and 47 836 false trials.

Since the conventional KL linear kernel has been proven to be superior to the GMM-UBM recognizer and polynomial expansion SVM system in many research reports [3], [8], [14], in this letter, we focus on comparing the proposed GUMI kernel [see (16)] with the conventional linear kernel [refer to (7)].

In the experiments, both systems strictly comply with the same training and testing conditions such as feature extraction processing, background databases for UBM and SVM, database for test-normalization (Tnorm) [13], and NAP training database. We carried out the experiment on gender-independent (GI) and gender-dependent (GD) modes, respectively. The databases used to train the speaker recognition systems are listed in Table I.

Authorized licensed use limited to: National University of Singapore. Downloaded on January 7, 2009 at 01:15 from IEEE Xplore. Restrictions apply.

TABLE I
CONFIGURATION OF TRAINING DATABASES

Training database	Gender-independence	Gender-dependence
UBM	Switchboard + SRE-2004 1side Train	SRE-2004 Train
NAP	SRE-2004 Train	SRE-2004 Train
SVM	Switchboard + SRE-2004 1side Train	SRE-2004 Train
Tnorm	SRE-2005 1conv. Train	SRE-2005 1conv. Train

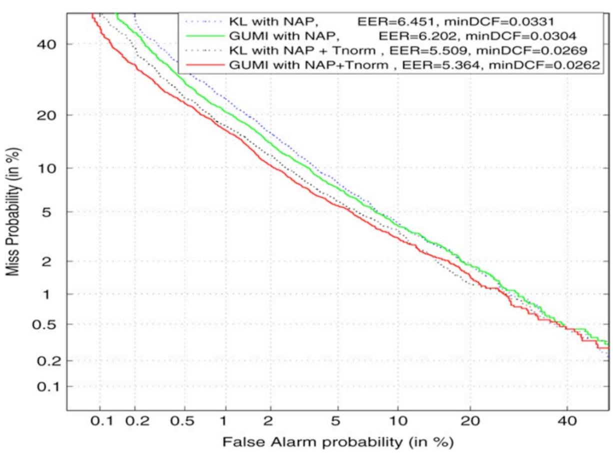

Fig. 2. DET curves of the GMM-supervector SVM systems with both KL and GUMI kernels, for gender-dependence testing mode for the 51448 trials (including male and female) for NIST SRE-2006 1conv4w-1conv4w.

TABLE II
COMPARISON BETWEEN THE GMM-SUPERVECTOR SVMS WITH THE KL KERNEL AND THE GUMI KERNEL BY USING THE 512-MIXTURES OF GMM ON THE SRE-2006 1CONV4W-1CONV4W EVALUATION, IN TERMS OF EQUAL ERROR RATE (EER) AND MINIMUM DETECTION COST FUNCTION (MINDCF)

System (51448 trials)	EER	minDCF$\times$100
KL kernel (GI)	11.16%	5.44
GUMI kernel (GI)	10.85%	5.19
KL kernel NAP (GI)	8.67%	4.14
GUMI kernel NAP (GI)	8.35%	3.19
KL kernel NAP (GD)	6.45%	3.31
GUMI kernel NAP (GD)	6.20%	3.04
KL kernel NAP +Tnorm (GD)	5.51%	2.69
GUMI kernel NAP +Tnorm (GD)	5.36%	2.62

For feature extraction, an 18-dimensional linear predictive cepstral coefficient (LPCC) vector is computed every 20 ms with 30-ms frame size. Delta-cepstral coefficients are obtained over a ± 2 frame span and appended to the cepstra producing a 36-dimensional feature vector. An energy-based speech detector is applied to discard silence and noise frames. To compose the GMM, 512 Gaussian mixture components are chosen. It leads to 18 432-dimension of the GMM-supervector. The GMM-supervector is obtained using an MAP adaptation in [1], where the relevance factor r is set to 16.

The NAP [11] is applied to the GMM-supervectors. The rank of the NAP projection matrix is selected to 80. The NIST SRE 2005 1conv4w training models are used as cohort models for Tnorm.

Fig. 2 plots the detection error trade-off (DET) curves of a gender-dependence testing. Table II gives the experimental results for the two kernels. It can be seen that both the figure and table have shown the consistent improvement by using the proposed kernel.

V. SUMMARY

In this letter, we propose the Bhattacharyya distance to measure the similarity between two GMM distributions. In GMM-supervector representation, the speaker cues from an utterance are transferred into a GMM through certain kind of transformation such as MAP adaptation. The speaker cues are statistically represented by the distribution parameters, e.g., mean, covariance, and mixture weight. Obviously, appropriate exploitation of more parameter improves the performance of the system. With the Bhattacharyya distance, we introduce a GMM-supervector named as GUMI supervector. The GUMI concept extends the kernel to exploit the information of covariance. Comparing with the conventional KL kernel, the proposed kernel reflects the salient characteristics of the speaker GMM by removing the common properties of the speakers which is reflected by UBM, and it more contains the inherent GMM information of the individual speaker characteristics. The validation of the proposed kernel is proven through experimentation on the core-test of the 2006 NIST SRE. The superiority of the new kernel is of persuadable proof and convinced reliability as compared to the KL kernel.

REFERENCES

[1] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digit. Signal Process.*, no. 10, pp. 19–41, 2000.

[2] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition," *Comput. Speech Lang.*, vol. 20, pp. 210–229, 2006.

[3] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, vol. 13, pp. 308–311, 2006.

[4] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.

[5] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. COM-15, pp. 52–60, Feb. 1967.

[6] W. M. Campbell, "A covariance kernel for SVM language recognition," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, 2008, pp. 4141–4144.

[7] M. Przybocki and A. Martin, The NIST Year 2006 Speaker Recognition Evaluation Plan, 2006. [Online]. Available: http://www.nist.gov/speech/tests/spk/2006/.

[8] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, France, 2006, pp. 97–100.

[9] J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis. Cambridge, U.K., 2004.

[10] R. Collobert and S. Bengio, "SVMTorch: Support vector machines for large-scale regression problems," *J. Mach. Learn. Res.*, vol. 1, pp. 143–160, 2001.

[11] A. Solomonoff, W. M. Campbell, and C. Quillen, "Channel compensation for SVM speaker recognition," in *Proc. Odyssey04*, 2004, pp. 57–62.

[12] M. Przybocki and A. Martin, "NIST speaker recognition evaluation chronicles," in *Proc. ODYSSEY, The Speaker and Language Recognition Workshop*, 2004.

[13] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *Digit. Signal Process.*, vol. 10, pp. 42–54, Jan. 2000.

[14] H. Hu, M. X. Xu, and W. Wu, "GMM supervector based SVM with spectral features for speech emotion recognition," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, 2007, vol. 4, pp. 413–416.

Authorized licensed use limited to: National University of Singapore. Downloaded on January 7, 2009 at 01:15 from IEEE Xplore. Restrictions apply.