INTERSPEECH 2009 BRIGHTON

Variational Dynamic Kernels for Speaker Verification

C. Longworth, R.C. van Dalen and M.J.F. Gales

Engineering Department, Cambridge University
Trumpington St, Cambridge, CB2 1PZ
{cl336,rcv25,mjfg}@eng.cam.ac.uk

10.21437/Interspeech.2009-388

Abstract

An important aspect of SVM-based speaker verification is the choice of dynamic kernel. Recently there has been interest in the use of kernels based on the Kullback-Leibler divergence between GMMs. Since this has no closed-form solution, typically a matched-pair upper bound is used instead. This places significant restrictions on the forms of model structure that may be used. All GMMs must contain the same number of components and must be adapted from a single background model. For many tasks this will not be optimal. In this paper, dynamic kernels are proposed based on alternative, variational approximations to the KL divergence. Unlike the matched-pair bound, these do not restrict the forms of GMM that may be used. Additionally, using a more accurate approximation of the divergence may lead to performance gains. Preliminary results using these kernels are presented on the NIST 2002 SRE dataset.

Index Terms: Speaker Verification, Support Vector Machines, Dynamic Kernels

1. Introduction

Speaker verification (SV) is a binary classification task in which the objective is to determine whether or not a speech utterance was spoken by a specific claimed speaker. There has been considerable interest and success in applying support vector machines (SVMs) to this task. Many state-of-the-art SV systems make use of distributional kernels, such as the GMM-supervector [1] and the nonlinear GMM-supervector kernels [2]. For these kernels, a generative model is trained to represent each utterance in the dataset. The kernel function between a pair of utterances is then derived from the Kullback-Leibler (KL) divergence between the corresponding models. For text-independent tasks, typically Gaussian mixture models (GMMs) are used. In this case there exists no closed-form solution of the KL divergence. Instead the *matched-pair* bound is used. However, this approximation requires that all GMMs contain the same number of components and that components with the same index are coordinated. In practice this means that all GMMs must be adapted from a single universal background model (UBM). This restriction may limit the performance of a SV system.

Recently, it has been shown that by introducing additional variational distributions over Gaussian components, more accurate approximations to the KL divergence can be derived. In this paper, two of these approximations, the variational approximation [3] and the variational upper bound [3, 4], are used to motivate new forms of dynamic kernel. Unlike the GMM-supervector kernel, these variational kernels do not restrict all GMMs to have the same structure. This allows more complex training schemes. For example, GMMs may be adapted from a range of gender or noise condition-dependent background models. Additionally, the use of an kernel that more accurately reflects the true KL divergence between GMMs may lead to gains. This paper is organised as follows: the next section introduces the KL divergence and describes two variational approximations to the divergence between GMMs. In section 3 the use of the KL divergence for SV is discussed and dynamic kernels are proposed based on the variational approximations. In section 4 preliminary experimental results on the NIST 2002 SRE task are presented. Finally conclusions are drawn.

Chris Longworth was funded by the Schiff Foundation. Rogier van Dalen was funded by Toshiba Research Europe Ltd.

2. KL divergence with GMMs

The Kullback-Leibler divergence defines the relative entropy between two distributions. For distributions f_i and f_j over $\boldsymbol{o}$, the divergence, $KL(f_i||f_j)$, is defined by

$$KL(f_i||f_j) = \int f_i(\boldsymbol{o}) \log \frac{f_i(\boldsymbol{o})}{f_j(\boldsymbol{o})} \, d\boldsymbol{o} \tag{1}$$

For Gaussian distributions, $\tilde{f}_i(\boldsymbol{o}) = \mathcal{N}(\boldsymbol{o}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ and $\tilde{f}_j(\boldsymbol{o}) = \mathcal{N}(\boldsymbol{o}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ the KL divergence can be expressed as

$$\begin{aligned} KL(\tilde{f}_i||\tilde{f}_j) &= \frac{1}{2}\big[\log \frac{|\boldsymbol{\Sigma}_j|}{|\boldsymbol{\Sigma}_i|} + Tr[{\boldsymbol{\Sigma}_j}^{-1}\boldsymbol{\Sigma}_i] - d \\ &\quad + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^{\mathsf{T}} {\boldsymbol{\Sigma}_j}^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\big] \end{aligned} \tag{2}$$

where d is the dimensionality of $\boldsymbol{o}$. For GMM distributions $f_i(\boldsymbol{o}) = \sum_{n=1}^{N} c_{in}\mathcal{N}(\boldsymbol{o}; \boldsymbol{\mu}_{in}, \boldsymbol{\Sigma}_{in})$ and $f_j(\boldsymbol{o}) = \sum_{m=1}^{M} c_{jm}\mathcal{N}(\boldsymbol{o}; \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm})$ there exists no closed-form solution. Instead the divergence must be estimated. This may be achieved either by using sampling approaches or by finding a closed-form expression that approximates the true KL divergence. Here the latter approach is examined. When $N = M$, a commonly used upper bound to the KL divergence is the *matched-pair* bound. This is defined by

$$KL^{\text{MP}}(f_i||f_j) = \sum_{m=1}^{M} c_{im} \left[\log \frac{c_{im}}{c_{jm}} + KL(f_{im}||f_{jm})\right] \tag{3}$$

where $KL(f_{in}||f_{jm})$ indicates the divergence between component n of f_i and component m of f_j. Permuting the components of one distribution will affect the obtained bound. Hence the matched-pair bound is only suitable when there is a clearly defined coordination between pairs of components from the two distributions. This may be the case when both are adapted from the same background distribution but will not be true generally. In the following subsections, two variational approximations are described that do not suffer from these restrictions.

2.1. Variational approximation

A variational approximation was derived by Hershey and Olsen [3]. Here the KL divergence is initially decomposed into the difference between two expected values

$$KL(f_i||f_j) = \int f_i(\boldsymbol{o}) \log f_i(\boldsymbol{o})\, d\boldsymbol{o} - \int f_i(\boldsymbol{o}) \log f_j(\boldsymbol{o})\, d\boldsymbol{o} \quad (4)$$

A lower bound can then be derived separately for each of the terms in 4. Starting with the second term, a bound may be obtained by introducing a discrete variational distribution $q_{m|n}$, such that $q_{m|n} > 0$ and $\sum_{m=1}^{M} q_{m|n} = 1$. Here, $f_{in}(\boldsymbol{o})$ is used to indicate the likelihood of $\boldsymbol{o}$ given the Gaussian distribution associated with component n of distribution f_i.

$$\begin{aligned}
&\int f_i(\boldsymbol{o}) \log f_j(\boldsymbol{o})\, d\boldsymbol{o} \\
&= \int \sum_{n=1}^{M} c_{in} f_{in}(\boldsymbol{o}) \log \sum_{m=1}^{M} q_{m|n} \frac{c_{jm} f_{jm}(\boldsymbol{o})}{q_{m|n}}\, d\boldsymbol{o} \\
&\geq \int \sum_{n=1}^{M} c_{in} f_{in}(\boldsymbol{o}) \sum_{m=1}^{M} q_{m|n} \log \frac{c_{jm} f_{jm}(\boldsymbol{o})}{q_{m|n}}\, d\boldsymbol{o} \quad (5)
\end{aligned}$$

This lower bound is tightest when the expression is maximised with respect to $q_{m|n}$. The optimal value $\hat{q}_{m|n}$ is given by

$$\hat{q}_{m|n} = \frac{c_{jm} e^{-KL(f_{in}||f_{jm})}}{\sum_{r=1}^{M} c_{jr} e^{-KL(f_{in}||f_{jr})}} \quad (6)$$

Similarly, a second variational distribution may be introduced to obtain a lower bound to the first term in equation 4. Taking the difference between the two lower bounds yields a variational approximation to the KL divergence.

$$KL^{\text{VAR}}(f_i||f_j) = \sum_{n=1}^{N} c_{in} \log \frac{\sum_{s=1}^{N} c_{is} e^{-KL(f_{in}||f_{is})}}{\sum_{m=1}^{M} c_{jm} e^{-KL(f_{in}||f_{jm})}} \quad (7)$$

Although equation 7 is not a strict bound, in practice it forms a close approximation to the KL divergence. This variational approximation is related to the matched-pair bound. For the case where f_i and f_j have equal numbers of components and when the variational distributions are only non-zero when $m = n$ the approximation becomes an upper bound and is equivalent to the form given in equation 3.

2.2. Variational upper bound

A variational upper bound was derived independently in [4] and [3]. Like the variational approximation, two discrete variational distributions $q_{m|n} \geq 0$ and $v_{n|m} \geq 0$ are introduced. Here, these distributions satisfy the following constraints $\sum_{m=1}^{M} q_{m|n} = c_{in}$ and $\sum_{n=1}^{N} v_{n|m} = c_{jm}$. Using Jensen's inequality the following bound may be obtained.

$$\begin{aligned}
KL(f_i||f_j) &= -\int f_i(\boldsymbol{o}) \log \sum_{m,n=1}^{M,N} \frac{q_{m|n} f_{in}(\boldsymbol{o})}{f_i(\boldsymbol{o})} \frac{v_{n|m} f_{jm}(\boldsymbol{o})}{q_{m|n} f_{in}(\boldsymbol{o})}\, d\boldsymbol{o} \\
&\leq -\int f_i(\boldsymbol{o}) \sum_{m,n=1}^{M,N} \frac{q_{m|n} f_{in}(\boldsymbol{o})}{f_i(\boldsymbol{o})} \log \frac{v_{n|m} f_{jm}(\boldsymbol{o})}{q_{m|n} f_{in}(\boldsymbol{o})}\, d\boldsymbol{o} \\
KL^{\text{UP}}(f_i||f_j) &= \sum_{m,n=1}^{M,N} q_{m|n} \left[\log \frac{q_{m|n}}{v_{n|m}} + KL(f_{in}||f_{jm}) \right] \quad (8)
\end{aligned}$$

This bound is tightest when q and v are selected to minimise equation 8. Unlike for the variational approximation there is no closed-form expression for the optimal q and v. However, by fixing one set of variational parameters and optimising the other the following update rules are obtained.

$$\begin{aligned}
{v_{n|m}}^{(k+1)} &= \frac{c_{jm} {q_{m|n}}^{(k)}}{\sum_{s=1}^{N} {q_{m|s}}^{(k)}} \quad (9) \\
{q_{m|n}}^{(k+1)} &= \frac{c_{in} {v_{n|m}}^{(k)} e^{-KL(f_{in}||f_{jm})}}{\sum_{r=1}^{M} {v_{n|r}}^{(k)} e^{-KL(f_{in}||f_{jr})}} \quad (10)
\end{aligned}$$

By iteratively reapplying equation 9 and 10 the upper bound will be tightened. In [3] the variational distributions were initialised to $v_{n|m} = q_{m|n} = c_{jm} c_{in}$ since any parameters that are set to zero will be unchanged after each iteration. Like the variational approximation, the variational upper bound is related to the matched-pair bound. When both distributions consist of the same number of components and $q_{m|n} = c_{in}$ and $v_{n|m} = c_{jm}$ where $m = n$ otherwise $q_{m|n} = v_{n|m} = 0$ the variational upper bound will have the same form as equation 3. The likelihood of this occurring in practice is related to the dimension d of the distribution. When d is high $q_{m|n}$ and $v_{n|m}$ are more likely to be sparse.

3. SVM-based Speaker Verification

SVMs have been successfully applied to a wide range of machine learning problems. One reason for this is that they can be kernelised. In SVM training and inference all references to data are in the form of inner-products between data examples. It is then possible to define a *kernel function* $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ that implicitly calculates the inner-product between two vectors in some, possibly very high dimensional, *feature space*. Standard forms of kernel, such as the polynomial or Gaussian kernel, have been found to provide gains over linear kernels on a range of tasks. One issue when applying SVMs to speech processing tasks is that most standard forms of kernel only operate on data of fixed dimensionality. However, speech utterances are typically variable length sequences $\boldsymbol{O} = \{\boldsymbol{o}_1, \ldots, \boldsymbol{o}_T\}$. This has led to the development of dynamic/sequence kernels. These kernels operate on sequences and have the form

$$K(\boldsymbol{O}_i, \boldsymbol{O}_j) = < \boldsymbol{\phi}(\boldsymbol{O}_i), \boldsymbol{\phi}(\boldsymbol{O}_j) > \quad (11)$$

Here $\boldsymbol{\phi}(\boldsymbol{O})$ is a function that maps a sequence of observations to a fixed dimensional vector. For parametric and derivative kernels [5] the form of $\boldsymbol{\phi}(\boldsymbol{O})$ is explicitly specified. Alternatively, $\boldsymbol{\phi}(\boldsymbol{O})$ may be implicitly defined by $K(\boldsymbol{O}_i, \boldsymbol{O}_j)$. The form of kernel function also defines the distance metric between two feature vectors. In the following sections, schemes are described for obtaining kernel functions suitable for SV from the approximations to the KL divergence introduced in section 2.

3.1. GMM-supervector kernels

The Kullback-Leibler divergence is commonly used as a similarity measure between distributions. It may therefore also be used to motivate kernel functions suitable for speaker verification. One approach is learn a distribution f_i associated with the speech from each utterance $\boldsymbol{O}_i$. A kernel function $K(\boldsymbol{O}_i, \boldsymbol{O}_j)$ is then defined between two utterances $\boldsymbol{O}_i$ and $\boldsymbol{O}_j$ based on an approximation to the KL divergence between f_i and f_j. For the GMM-supervector [1] and related kernels, f_i and f_j are constrained to be GMMs that differ only in the means. Under these conditions, the matched-pair bound, defined in equation 3, may

be used. Since the KL divergence is asymmetric, the symmetric KL divergence, $KL(f_i||f_j) + KL(f_j||f_i)$, is often used instead. This equals zero if and only if $f_i = f_j$, otherwise it is positive. These properties are typical of a distance metric rather than an inner product. It is therefore common to alter the function prior to use such that it behaves more like an inner product. One approach is to make use of the polarisation identity, $D(a,b)^2 = K(a,a) - 2K(a,b) + K(b,b)$. This defines a relationship between a distance $D(a,b)$ and a kernel function $K(a,b)$ in a particular space. For the GMM-supervector kernel, the distance between f_i and f_j is defined by

$$D(f_i, f_j)^2 = KL^{\text{MP}}(f_i||f_j) + KL^{\text{MP}}(f_j||f_i) \tag{12}$$

This distance is related, via the polarisation identity, to the standard GMM-supervector [1] kernel function.

$$K^{\text{GMM-SV}}(\boldsymbol{O}_i, \boldsymbol{O}_j) = \sum_{m=1}^{M} c_m \boldsymbol{\mu}_{im}{}^{\text{T}} \boldsymbol{\Sigma}_m{}^{-1} \boldsymbol{\mu}_{jm} \tag{13}$$

An alternative approach to obtaining a kernel function from a distance metric is via exponentiation. When $KL(f_i||f_j)$ is approximated using the matched-pair bound this yields

$$K^{\text{MP}}(\boldsymbol{O}_i, \boldsymbol{O}_j) = e^{-\alpha \sum_{m=1}^{M} c_m (\boldsymbol{\mu}_{im} - \boldsymbol{\mu}_{jm})^{\text{T}} \boldsymbol{\Sigma}_m{}^{-1} (\boldsymbol{\mu}_{im} - \boldsymbol{\mu}_{jm})} \tag{14}$$

where α is a constant scaling term. Equation 14 has the form of the nonlinear GMM-supervector kernel used in [2]. A related kernel, based on Gaussian distributions, was used in [6]. Polarisation and exponentiation are closely related. When $\alpha = 1/2\sigma^2$ equation 14 is equivalent to applying an RBF kernel to standard GMM-supervector features. However, unlike the GMM-supervector kernel, $K^{\text{MP}}(\boldsymbol{O}_i, \boldsymbol{O}_j)$ does not have an explicit associated feature-space.

3.2. Variational dynamic kernels

For kernels that are derived using the matched-pair bound, such as (13) and (14), any experimental conditions that weaken the coordination between components are likely to also degrade performance. This includes performing multiple iterations of adaptation or applying MAP with low values of τ. Alternatively, kernels may be derived using either the variational approximation or the variational upper bound introduced in section 2. Due to the form of these approximations it is difficult to obtain a suitable kernel using the polarisation identity. This will be examined in further work. Here, exponentiation is used in a similar approach to equation 14. For the variational approximation, the following kernel is obtained.

$$K^{\text{VAR}}(\boldsymbol{O}_i, \boldsymbol{O}_j) = e^{-\alpha \left[KL^{\text{VAR}}(f_i||f_j) + KL^{\text{VAR}}(f_j||f_i) \right]} \tag{15}$$

A kernel may also be derived using the variational upper bound.

$$K^{\text{UP}}(\boldsymbol{O}_i, \boldsymbol{O}_j) = e^{-\alpha \left[KL^{\text{UP}}(f_i||f_j) + KL^{\text{UP}}(f_j||f_i) \right]} \tag{16}$$

Evaluating equation 16 requires optimising four sets of variational parameters. (This process is implicit for the variational approximation.) When both GMMs consist of M components and the variational parameters between components n and m are non-zero only when $n = m$, the obtained kernels will be identical to equation 14. Unlike the GMM-supervector or nonlinear GMM-supervector kernels, there is no requirement that all GMMs have the same structure or are adapted from the same background model. There are a number of situations where this is useful. If the duration of utterances vary greatly within the dataset, gains may be obtained by allowing the number of components per GMM to vary. Hence an utterance dependent-tradeoff could be made between increasing model flexibility and avoiding overfitting the data. Alternatively, when utterances come from speakers of varying genders or dialects, or are recorded under a range of different noise-conditions it may be advantageous to adapt each utterance from a background model that more closely resembles the characteristics of the utterance. This approach may also be combined with speaker-clustering schemes to obtain more accurate background models.

4. Experimental Results

The variational dynamic kernels were evaluated on the 2002 NIST SRE one-speaker detection task [7][1]. Each utterance was parameterised using a frame rate of 10ms and a window size of 30ms. 31 features were extracted per frame, these consisted of 15 static, 15 delta Mel-PLP coefficients and the delta energy. Cepstral Feature Warping was performed on each utterance using a three second window to introduce additional robustness to channel noise. 512-component, diagonal-covariance, gender-dependent UBMs were trained by EM using all enrollment utterances of the appropriate gender. For each training and test speech utterance, a corresponding GMM distribution was adapted from the appropriate background model. Two iterations of static-prior mean-only MAP were applied with $\tau = 1$.

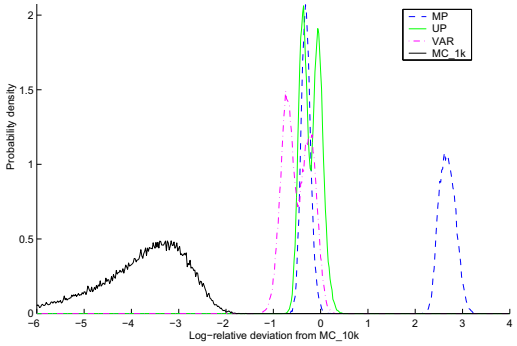

Figure 1: Log-relative deviation between KL-approximation and `MC_10k` over all pairs of enrollment utterances.

Initially, the different KL divergence approximations were compared. Figure 1 shows the distribution of the log-relative deviation between various approximation scheme and a reference estimate of the true KL divergence estimated over all pairs of enrollment utterances. This estimate was obtained using the Monte Carlo approach described in [3] using 10,000 samples. The difference in accuracy between using 10,000 (`MC_10k`) and 1,000 (`MC_1k`) independently drawn samples is shown in Figure 1. This was small indicating that `MC_10k` is a reasonable estimate of the true divergence. The matched-pair bound, variational approximation and variational upper bound were calculated using equations 3, 7 and 8 respectively. For the variational upper bound, $q_{m|n}$ and $v_{n|m}$ were initialised to $q_{m|n} = v_{n|m} = c_{in} c_{jm}$ and re-estimated using equations 9 and 10 for 15 iterations. For within-gender comparisons, there was a strong coordination between pairs of Gaussian components. Here the accuracy of the three approximations was similar. For

[1] The 2002 SRE data was chosen as it is one of the most recent evaluation datasets to be made generally available through the LDC. Later datasets are currently only available to SRE participants. However the techniques discussed here may be easily applied to more recent tasks.

the variational upper bound the optimal variational parameters were sparse and consistently approached the matched-pair solution. This was true to a lesser extent for the variational approximation which generally provided a closer approximation to the true KL divergence. For cross-gender comparisons, the true KL divergence was approximately twice that of within-gender comparisons. This was due to the use of gender-dependent UBMs. In these cases there was no clear coordination between components. Here the matched-pair bound typically exceeded the true KL divergence by an order of magnitude. In comparison, for the two variational approximations cross and within-gender accuracy was similar.

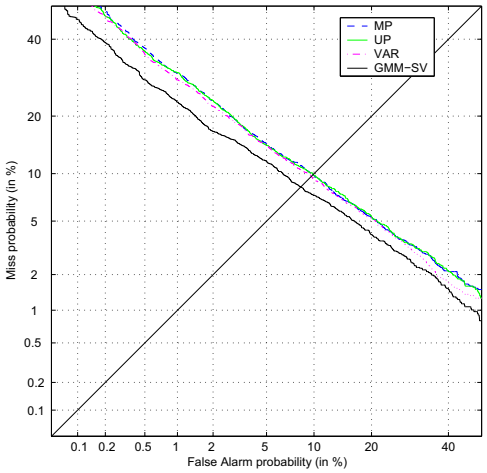

Figure 2: Gender-dependent system using various kernels

Next, the variational kernels were evaluated. SVM classifiers were trained using SVM^{light} [8] with C set to the default value. Initially, imposter examples were obtained from the enrollment data associated with other speakers of the same gender[2]. Kernels were evaluated based on the matched-pair bound (`MP`), variational (`VAR`) and variational upper bound (`UP`) approximations. For all exponential kernels α was fixed at 1. During preliminary experiments this optimised performance of `MP`. A GMM-supervector (`GMM-SV`) kernel was also evaluated. The DET curve in Figure 2 compares the performance of all kernels. `GMM-SV` performance was significantly better than the other kernels despite using the same GMM models. This agrees with results reported in [2]. There, gains were achieved by normalising the matched-pair estimate of the divergence between the background model and each GMM. This is nontrivial for the variational approximations and will be the subject of future work. Overall the three kernels based on exponential KL-approximations performed at a roughly similar level. The best performing kernel was (`VAR`) with an EER of 9.68%. Since the experimental setup did not require cross-gender kernel evaluations these results are in line with the accuracies in Figure 1.

Finally, the kernels were evaluated using GMMs adapted from a range of background models. Here the imposter set consisted of enrollment speech from both genders. Results for this experimental setup, (`GD-UBM`), are shown in Table 1 and compared with the strictly gender-dependent setup (`GD`) used in Figure 2. For all systems, minDCF results were in line with reported EERs. For the matched-pair kernel, including cross-speaker imposter data degraded performance by 0.42% EER.

[2]The setup used did not conform to the NIST SRE protocol, since enrollment data was used for both UBM training and imposter modelling. This was necessary due to the limited amount of development data available to the authors.

System	Equal Error Rate (%)			
	`GMM-SV`	`VAR`	`UP`	`MP`
`GD`	8.35	9.68	9.89	9.91
`GD-UBM`	8.38	9.59	9.69	10.29
`GI`	8.89	9.91	10.05	10.11

Table 1: Kernel performance using a) gender-dependent UBMs and imposter data (`GD`), b) gender-dependent UBMs and gender-independent imposter data (`GD-UBM`) and c) gender-independent UBM and imposter data (`GI`)

The fact that this loss is relatively small, and does not occur for the `GMM-SV` system, is due to the ability of the SVM to select appropriate support vectors. For `MP`, only 7% of imposter support vectors came from cross-gender speakers. For `VAR` and `UP`, small performance gains were observed. To establish whether this was simply due to the additional imposter data, a third experimental setup (`GI`) was also evaluated. Here all distributions were adapted from a single gender-independent background model. Again, imposter data from both genders was used for each target speaker. For all kernels evaluated this system performed worst. This degradation was primarily because the utterance-dependent distributions were less well adapted to the speech. Unfortunately, the lack of cross-gender trials in the 2002 NIST SRE meant that the gains that may be obtained by including cross-gender imposter data are limited.

5. Conclusion

This paper has described how kernel functions suitable for speaker verification may be derived from two variational approximations to the KL divergence between GMMs. Unlike standard dynamic kernels such as the GMM-supervector that are based on the matched-pair bound, these may be applied when models vary in structure or are adapted from different background distributions. Preliminary experimental results were presented on the NIST 2002 task. Future work will examine the use of more complex training schemes, for example where GMM-structure is allowed to vary, or where clusters of speakers are adapted from a wider range of background models. The use of KL divergence based model normalisation, as in [2], or development of variational kernels derived using polarisation may also lead to further gains.

6. References

[1] W. Campbell, D. Sturim, D. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation," in *ICASSP*, 2006.

[2] R. Dehak, N. Dehak, P. Kenny, and P. Dumouchel, "Linear and non linear GMM supervector machines for speaker verification," in *Proc. ICSLP*, 2007.

[3] J. Hershey and P. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *ICASSP*, 2007.

[4] K. Yu, "Adaptive training for large vocabulary continuous speech recognition," Ph.D. dissertation, University of Cambridge, 2006.

[5] C. Longworth and M. J. F. Gales, "Derivative and parametric kernels for speaker verification," in *Proc. ICSLP*, 2007.

[6] P. Moreno, P. Ho, and B. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications," in *Proc. Advances in NIPS*, 2004.

[7] A. Martin, "The NIST year 2002 speaker recognition evaluation plan," 2002, available from http://www.itl.nist.gov/iad/mig/tests/sre/2002.

[8] T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods - Support Vector Learning*, B. Scholkopf, C. Burges and A. Smola, Ed. MIT Press, 1999.