ALLEVIATING THE SMALL SAMPLE-SIZE PROBLEM IN I-VECTOR BASED SPEAKER VERIFICATION

Wei RAO and Man-Wai MAK

Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University, Hong Kong SAR, China
ellen.wei-rao@connect.polyu.hk, enmwmak@polyu.edu.hk

ABSTRACT

This paper investigates the small sample-size problem in i-vector based speaker verification systems. The idea of i-vectors is to represent the characteristics of speakers in the factors of a factor analyzer. Because the factor loading matrix defines the possible speaker- and channel-variability of i-vectors, it is important to suppress the unwanted channel variability. Linear discriminant analysis (LDA), within-class covariance normalization (WCCN), and probabilistic LDA are commonly used for such purpose. These methods, however, require training data comprising many speakers each providing sufficient recording sessions for good performance. Performance will suffer when the number of speakers and/or number of sessions per speaker are too small. This paper compares four approaches to addressing this small sample-size problem: (1) preprocessing the i-vectors by PCA before applying LDA (PCA+LDA), (2) replacing the matrix inverse in LDA by pseudo-inverse, (3) applying multi-way LDA by exploiting the microphone and speaker labels of the training data, and (4) increasing the matrix rank in LDA by generating more i-vectors using utterance partitioning. Results based on NIST 2010 SRE suggests that utterance partitioning performs the best, followed by multi-way LDA and PCA+LDA.

Index Terms— Speaker verification, i-vectors, LDA, utterance partitioning, multi-way LDA.

1. INTRODUCTION

Current state-of-the-art speaker verification systems typically represent the acoustic characteristics of a speaker by converting his/her variable-length utterances into fixed-length vectors. These vectors, called identity vectors or i-vectors for short [1], live on a low-dimensional space known as the total variability space. Given a training set containing utterances produced by many speakers, the total variability space can be obtained by factor analysis [1, 2]. This space represents the possible variability, including speaker and channel variability, of i-vectors. Specifically, the GMM-supervector [3] representation of an utterance is given by[1]

$$\boldsymbol{m}_s = \boldsymbol{m} + \boldsymbol{T}\mathbf{w}_s \tag{1}$$

where $\boldsymbol{m}$ is the GMM-supervector representation of the universal background model (UBM), $\boldsymbol{T}$ is a low-rank loading matrix representing the total variability space, and $\mathbf{w}_s$ is the i-vector that comprises the factors (latent variables) representing the speaker's characteristics. Therefore, instead of using the high-dimensional supervector $\boldsymbol{m}_s$ to represent the speaker s, the i-vector approach represents a speaker by a low-dimension vector $\mathbf{w}_s$, typically of dimension 400. During verification, given a test utterance, the i-vector $\mathbf{w}_t$ corresponding to the test utterance is compared with the i-vector of the claimed identity using the cosine distance measure.

Representing the i-vectors in a low-dimensional space opens up opportunity for using machine learning techniques such as linear discriminant analysis (LDA) [4], within-class covariance normalization (WCCN) [5] and probabilistic LDA (PLDA) [6] to suppress session- and channel-variability. The key idea is to estimate the channel variability from these training data and to project the target and test i-vectors to a subspace with minimum channel variability. While these techniques have achieved state-of-the-art performance in recent NIST Speaker Recognition Evaluations (SRE), they require multiple training speakers each providing sufficient numbers of sessions to train the transformation matrices. When the number of training speakers and/or number of recording sessions per speaker are insufficient, numerical difficulty or error will occur in estimating the transformation matrices, resulting in inferior performance. This paper investigates and proposes several approaches to overcoming this resource-constrained scenario:

1. *PCA+LDA*. The numerical difficulty in estimating the transformation matrices is due to insufficient rank in the within-speaker covariance matrix. We investigated using PCA to project the training i-vectors to a lower dimension space prior to computing the within-speaker scatter matrix [7, 8]. With the reduction in the dimension of i-vectors, the rank requirement of LDA and WCCN can be reduced to a comfortable level for reliable estimation of the LDA and WCCN transformation matrices.
2. *Pseudo-inverse LDA*. The rank deficiency problem can be avoided by replacing the inverse of the within-speaker scatter matrix by its pseudo inverse [9, 10]. The idea is that during eigen-decomposition, any components with eigenvalues smaller than a threshold will be automatically discarded by the pseudo-inverse procedure.
3. *Multi-way LDA*. In the classical i-vector based approach, covariance analysis is only applied to the speaker domain for computing the within-speaker scatter matrix and between-speaker scatter matrix. The assumption is that each training i-vector has a speaker label and that each speaker provides a number of utterances (i-vectors) using a variety of microphones. However, the approach ignores the fact that in most cases the same set of microphones are used in the recording sessions for all training speakers. We propose exploiting this extra information to strengthen the discriminative capability

[1] A GMM-supervector is formed by stacking the mean vectors of a Gaussian mixture model.

of LDA (see Section 4 for details). We refer to this approach as "Multi-way LDA".

4. *Utterance Partitioning*. We applied our previously proposed utterance partitioning technique [11, 12] to create more sessions and i-vectors per training speaker to estimate the transformation matrices. More precisely, rather than using a single i-vector to represent a full-length utterance, the utterance is partitioned into a number of sub-utterances whose length is long enough for the i-vectors to capture the speaker characteristics. For example, if a full-length utterance is divided into four sub-utterance, a total of five i-vectors can be obtained.

Our key findings are that when the number of sessions per speaker for training the LDA and WCCN projection matrices is less than four, both PCA and pseudo-inverse can help alleviate the numerical difficulty occurred in estimating the inverse of the within-speaker scatter matrices. It was found that multi-way LDA can make better use of the structured information in the training i-vectors, thus resulting in better performance than both PCA+LDA and pseudo-inverse LDA. The best performance is achieved by utterance partitioning. The reason is that, unlike the other methods, utterance partitioning can make the full use of the limited training data by avoiding the information contents of i-vectors to become saturated [12].

2. RELATED WORK ON SMALL SAMPLE-SIZE PROBLEMS

There are many applications in which the dimensionality of data is larger than the number of training samples. For examples, in microarray data analysis [13], the number of genes tends to be much larger than the number of samples. In face recognition [14], the feature dimension is usually very high because it is proportional to the number of pixels. In machine learning literature, this is known as the small sample-size problem.

To apply LDA for classification or dimension reduction, the small sample-size problem will become an issue when the number of training samples is small but the feature dimension is high. This is because the LDA solution requires the computation of the inverse of the within-class scatter matrix, which may become singular when the number of training samples is small. Over the years, a number of methods have been proposed to address this problem.

A simple approach is to use PCA to reduce the dimension of the original vector before applying LDA to find the optimal discriminant subspace [7, 8]. The method is known as PCA+LDA in the literature. The dimension of the PCA projection space is selected such that the within-class scatter matrix becomes nonsingular so that LDA can be applied without numerical difficulty. The singularity problem has also been overcome by replacing the matrix inverse by pseudo-inverse [9, 10] or by adding a constant to the diagonal elements of the scatter matrix [15]. The former is called pseudo-inverse LDA and the latter is known as regularized LDA. The advantage of pseudo-inverse LDA is that components with eigenvalues smaller than a threshold are automatically discarded, thus avoiding the singularity problem. While it can be shown that the regularized scatter matrix is always nonsingular, the regularized LDA is harder to use because the amount of diagonal offset needs to be determined by cross validation. A more general formed of regularized LDA is the penalized LDA [16]. Instead of adding a positive diagonal matrix to the scatter matrix, penalized LDA adds a symmetric and positive semi-definite matrix to the scatter matrix in order to produce spatially smooth LDA coefficients.

More recently, null-space LDA [17, 18] and orthogonal LDA [19, 20] have been proposed to address the small sample-size problem. In null-space LDA, the between-class distance is maximized in the null space of the within-class scatter matrix. The singularity problem is implicitly avoided because no matrix inverse is needed. The method reduces to the conventional LDA when the within-class scatter matrix has full rank. In orthogonal LDA, the discriminant vectors are orthogonal to each other, and the optimal transformation matrix is obtained by simultaneous diagonalization of the between-class, within-class, and total scatter matrices. Again, the singularity problem has been avoided because matrix inverse is only applied to the diagonal matrix containing non-zero singular values [19]. It has been shown that when the rank of the total scatter matrix is equal to the sum of the rank of the between-class and within-class scatter matrices, orthogonal LDA is equivalent to null-space LDA [20].

3. I-VECTOR BASED SPEAKER VERIFICATION

Because i-vectors contain both speaker and channel variation in the total variability space, inter-session compensation plays an important role in the i-vector framework. It was found in [1] that projecting the i-vectors by LDA followed by WCCN achieves the best performance.

The idea of LDA is to find a set of orthogonal axes for minimizing the within-class variation and maximizing the between-class separation. In the i-vector framework, the i-vectors of a speaker constitute a class, leading to the following objective function [4]:

$$\mathbf{A} = \underset{\mathbf{A}}{\operatorname{argmax}} \left\{ \operatorname{tr} \left[\left(\mathbf{A}^{\mathsf{T}} \mathbf{S}_{ws} \mathbf{A} \right)^{-1} \left(\mathbf{A}^{\mathsf{T}} \mathbf{S}_{bs} \mathbf{A} \right) \right] \right\} \tag{2}$$

where $\mathbf{A}$ defines the optimal discriminant subspace on which the i-vectors should be projected, $\mathbf{S}_{ws}$ is the within-speaker scatter matrix, and $\mathbf{S}_{bs}$ is the between-class scatter matrix. Given a set of training i-vectors $\{\mathbf{w}_j^i; i = 1, \ldots, S, j = 1, \ldots, M_i\}$ where S is the number of training speakers and M_i is the number of utterances from the i-th training speaker, these two scatter matrices are written as:

$$\mathbf{S}_{ws} = \sum_{i=1}^{S} \frac{1}{M_i} \sum_{j=1}^{M_i} (\mathbf{w}_j^i - \boldsymbol{\mu}^i)(\mathbf{w}_j^i - \boldsymbol{\mu}^i)^{\mathsf{T}} \tag{3}$$

and

$$\mathbf{S}_{bs} = \sum_{i=1}^{S} (\boldsymbol{\mu}^i - \boldsymbol{\mu})(\boldsymbol{\mu}^i - \boldsymbol{\mu})^{\mathsf{T}}, \tag{4}$$

where $\boldsymbol{\mu}^i = \frac{1}{M_i} \sum_{j=1}^{M_i} \mathbf{w}_j^i$ is the mean i-vector of the i-th speaker and $\boldsymbol{\mu}$ is the global mean of all i-vectors in the training dataset. Maximizing Eq. 2 leads to the projection matrix $\mathbf{A}$ that comprises the leading eigenvectors of $\mathbf{S}_{ws}^{-1}\mathbf{S}_{bs}$.

WCCN [5] was originally used for normalizing the kernels in SVMs. In the i-vector framework, WCCN is to normalize the within-speaker variation. Dehak et al. [1] found that the best approach is to project the LDA reduced i-vectors to a subspace specified by the square-root of the inverse of the following within-class covariance matrix:

$$\mathbf{W} = \sum_{i=1}^{S} \frac{1}{M_i} \sum_{j=1}^{M_i} (\mathbf{A}^{\mathsf{T}} \mathbf{w}_j^i - \widetilde{\boldsymbol{\mu}^i})(\mathbf{A}^{\mathsf{T}} \mathbf{w}_j^i - \widetilde{\boldsymbol{\mu}^i})^{\mathsf{T}} \tag{5}$$

where $\widetilde{\boldsymbol{\mu}^i} = \frac{1}{M_i} \sum_{j=1}^{M_i} \mathbf{A}^{\mathsf{T}} \mathbf{w}_j^i$ and $\mathbf{A}$ is the LDA projection matrix. The WCCN projection matrix $\mathbf{B}$ can be obtained by Cholesky decomposition of $\mathbf{W}^{-1} = \mathbf{B}\mathbf{B}^{\mathsf{T}}$.

During verification, the cosine distance between the claimant's i-vector ($\mathbf{w}_t$) and target-speaker's i-vector ($\mathbf{w}_s$) in the LDA+WCCN projection space [21]:

$$S_{\cos}(\mathbf{w}_t, \mathbf{w}_s) = \frac{\left\langle \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{w}_t, \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{w}_s \right\rangle}{\left\|\mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{w}_t\right\| \left\|\mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{w}_s\right\|}. \quad (6)$$

The score is then further normalized (typically by ZT-norm [22]) before comparing with a threshold for making a decision.

4. MULTI-WAY LINEAR DISCRIMINANT ANALYSIS

Conventional LDA uses the information of speaker labels and a variety of microphone recordings per speaker to obtain the within-speaker and between-speaker scatter matrices. As a result, the method performs covariance analysis on the speaker domain only, ignoring the fact that the training speakers typically use the same set of microphones for recording. Here, we propose exploiting this extra information to strengthen the discriminative capability of LDA.

More precisely, the i-vectors of the training speakers are arranged in a grid, where the rows represent the speakers, the columns represents the microphones, and each element in the grid represents an i-vector. The dimension of i-vectors is firstly reduced by projecting the i-vectors to a subspace that maximizes the within-microphone variation, which represents the dispersion of i-vectors along the columns of the grid. The objective function is:

$$\mathbf{C} = \underset{\mathbf{C}:\|\mathbf{c}_i\|=1}{\operatorname{argmax}} \left[\operatorname{tr}\left(\mathbf{C}^{\mathsf{T}}\mathbf{S}_{wm}\mathbf{C}\right)\right] \quad i = 1, \ldots, L \quad (7)$$

where $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \cdots\ \mathbf{c}_L]$ defines the optimal discriminant subspace of dimension L on which the i-vectors should be projected and $\mathbf{S}_{wm}$ is the within-microphone scatter matrix:

$$\mathbf{S}_{wm} = \frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{S}(\mathbf{w}_j^i - \boldsymbol{\mu}_j)(\mathbf{w}_j^i - \boldsymbol{\mu}_j)^{\mathsf{T}} \quad (8)$$

where $\boldsymbol{\mu}_j = \frac{1}{S}\sum_{i=1}^{S}\mathbf{w}_j^i$ is the mean i-vector of the j-th microphone, S is the number of training speakers, and M is the number of microphones. Maximizing Eq 7 leads to the projection matrix $\mathbf{C}$ that comprises the L leading eigenvectors of $\mathbf{S}_{wm}$. Then, conventional LDA can be applied to the dimension reduced i-vectors, which amounts to finding a subspace that maximize the speaker separability but minimize the within speaker variability (along the rows of the grid).

Note that unlike PCA where the labels of training data are ignored, Eqs. 7 and 8 make use of both speaker and microphone labels of the training data. The use of microphone labels is expected to find a more discriminative subspace than the one found by PCA. Because for each column in the grid, the i-vectors are produced by different speakers using the same microphone, more discriminative subspace can be found by maximizing the separability of different speakers using the same microphone. It is however not desirable to minimize the between-microphone variability because the rank of between-microphone scatter matrix $\mathbf{S}_{bm}$ is typically very small. For example, in our experiments, the maximum value of M is 8, meaning that the rank of $\mathbf{S}_{bm}$ is only 7.

5. EXPERIMENTS

Speech Data and Acoustic Features: The *extended core set* of NIST 2010 Speaker Recognition Evaluation (SRE) was used for performance evaluation. This paper focuses on the interview and microphone speech of the extended core task, i.e., Common Conditions 1, 2, 4, 7 and 9. The equal error rate (EER) and the new minimum Detection Cost Function (DCF) were used as performance indicators. NIST 2005–2008 SREs were used as development data (UBM, total variability subspace training, LDA, WCCN, T-norm, and ZT-norm). Only the interview and microphone speech of male speakers in these corpora were used. Silence regions of the utterances in these corpora were removed by a VAD [23]. Cepstral mean normalization [24] was then applied to the MFCCs, followed by feature warping [25] using a window of 3 seconds. 19 MFCCs together with energy plus their 1st- and 2nd-derivatives were extracted from the speech regions of each utterance, leading to 60-dim acoustic vectors.

Total Variability Modeling and Channel Compensation: The i-vector systems use a gender-dependent UBM with 1024 mixtures. We selected 6,102 utterances from 191 speakers (each with at least 8 utterances) in NIST 2005–2008 SRE to estimate a total variability matrix with 400 total factors. A modified version of the BUT JFA Matlab code was used for i-vector training and scoring. Before calculating the verification scores, LDA and WCCN projections were performed for channel compensation. We used the same data set for training the total variability matrix to estimate the LDA and WCCN matrices. After LDA and WCCN projections, the dimension of i-vectors was reduced to 150.

Scoring Method and Score Normalization: We adopted cosine distance scoring. ZT-norm [22] was used for score normalization. 288 T-norm utterances and 288 Z-norm utterances (each from a different set of speakers) were selected from the interview and microphone speech in NIST 2005–08 SREs.

6. RESULTS AND DISCUSSIONS

6.1. Comparison of Different Methods

This experiment aims to investigate the performance of three different methods for solving the small sample-size problem. These methods are PCA + LDA, pseudo-inverse LDA, and utterance partitioning. Table 1 shows the performance achieved by these approaches when the number of recording sessions per training speaker (M) increases from 2 to 8 or above. The performance is obtained by concatenating the scores under Common Conditions 1, 2, 4, 7, and 9 in NIST 2010 SRE. The performance achieved by "Without LDA and WCCN" is considered as the baseline. For "LDA+WCCN", the performance is very poor when $M \leq 3$, because the within-speaker scatter matrix is close to singular. Only when $M \geq 4$, the benefit of LDA+WCCN becomes apparent. These observations also agree with the findings in [26].

Table 1 also shows the following properties:

1. when $M \leq 3$, pseudo-inverse LDA can help avoid the singularity problem. However, this methods lead to i-vectors that perform even poorer than those without LDA+WCCN projections. When the within-class scatter matrices have full rank ($M \geq 4$), the performance of pseudo-inverse LDA is the same as the classical LDA.
2. Preprocessing the i-vectors by PCA can not only avoid the singularity problem but also help the LDA to find a better projection matrix. However, when the rank of within-class scatter matrices is too low (e.g., when $M = 2$), the performance of PCA is poorer than those without LDA+WCCN projections. Moreover, the effect of PCA diminishes when the number of recordings per training speaker is sufficient ($M \geq 8$).
3. Utterance partitioning is an effective way to produce more informative i-vectors from a single utterance, thus effectively

Systems	No. of utts. per speaker (M)						
	2	3	4	5	6	7	≥ 8
(A) Without LDA and WCCN	12.60	12.60	12.60	12.60	12.60	12.60	12.60
(B) LDA + WCCN	23.39	22.25	6.98	5.51	4.59	4.22	2.98
(C) PI-LDA + WCCN	19.02	20.90	6.98	5.51	4.59	4.22	2.98
(D) PCA + LDA + WCCN	13.37	9.05	6.29	5.14	4.32	3.86	2.98
(E) UP-AVR + LDA + WCCN	**6.14**	**5.08**	**4.46**	**3.88**	**3.83**	**3.65**	**2.90**

(a) EER(%)

Systems	No. of utts. per speaker (M)						
	2	3	4	5	6	7	≥ 8
(A) Without LDA and WCCN	**0.90**	0.90	0.90	0.90	0.90	0.90	0.90
(B) LDA + WCCN	1.00	1.00	0.87	0.81	0.76	0.75	**0.63**
(C) PI-LDA + WCCN	0.99	1.00	0.87	0.81	0.76	0.75	0.63
(D) PCA + LDA + WCCN	1.00	0.95	0.88	0.82	0.77	**0.73**	0.63
(E) UP-AVR + LDA + WCCN	0.91	**0.87**	**0.82**	**0.78**	**0.75**	0.74	0.65

(b) MinNDCF

Table 1: The performance of different methods for alleviating the small sample-size problem in LDA. $M = x$ means each speaker only has x recordings for training the LDA and WCCN matrices. $M \geq 8$ means each speaker provides at least 8 recordings, with an average of 31 recordings per speaker. *"LDA"*: the conventional LDA; *"PI-LDA"*: pseudo-inverse LDA; *"PCA + LDA"*: perform PCA before LDA.

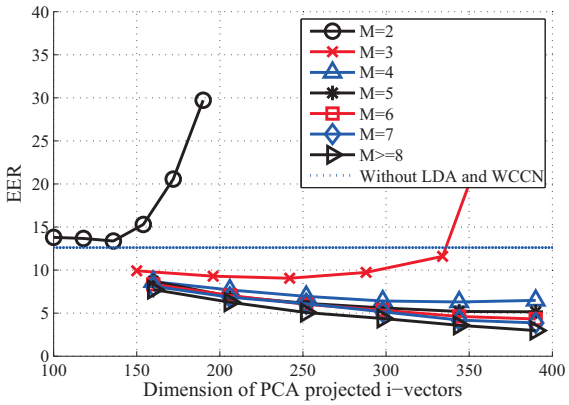

Fig. 1: EER versus the dimension after PCA projection. $M = x$ means each speaker only has x recordings for training the LDA and WCCN matrices.

avoiding the singularity problem in LDA. It also achieves the best performance among all methods investigated.

Fig. 1 shows the effect of varying the dimension of PCA projection on the performance of PCA+LDA. The results suggest that when the number of sessions per speaker (M) is equal to two, PCA cannot help the LDA for all projection dimension. In fact, the performance is even poorer than that without LDA (dotted line). This is caused by insufficient data for training the LDA, even though PCA can alleviate the singularity problem. The result also suggest that setting the PCA projection dimension close to the rank of within-class scatter matrices is not a good idea when $M \leq 3$.

6.2. Multi-way Linear Discriminant Analysis

To compare the effectiveness of PCA+LDA and multi-way LDA, we selected 63 male speakers from NIST 2008 SRE for training the LDA and WCCN projection matrices. Unlike the previous experiments, these speakers use the same set of microphones in the recording sessions. This arrangement allows us to arrange the training i-vectors in a grid, as explained in Section 4.

Systems	MinNDCF		EER (%)	
	$M = 7$	$M = 8$	$M = 7$	$M = 8$
(A) Without LDA and WCCN	0.90	0.90	12.60	12.60
(B) LDA + WCCN	–	0.99	–	15.29
(C) PI-LDA + WCCN	1.00	0.99	23.33	15.29
(D) PCA + LDA + WCCN	0.97	0.96	10.27	9.13
(E) MW-LDA + WCCN	**0.92**	**0.93**	**9.97**	**8.85**

Table 2: The performance of Multi-way LDA and other LDA methods. *MW-LDA + LDA*: Multi-way LDA. $M = x$ means each speaker only has x recordings for training the LDA and WCCN matrices. "–" denotes the situation where singularity occurs when estimating the projection matrices.

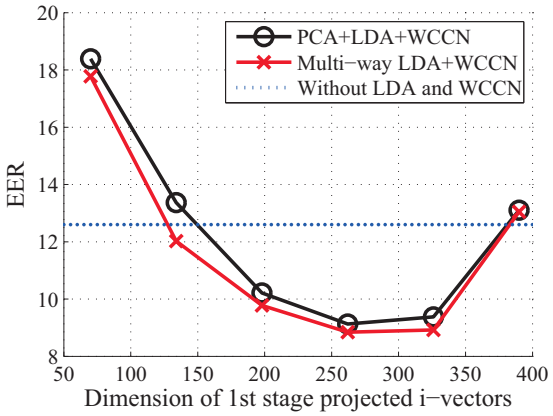

Fig. 2: EER versus the dimension of the projected i-vectors in the first stage of PCA+LDA and multi-way LDA. The number of recordings per speaker is 8 ($M = 8$). Refer to Section 6 for the explanation of "1st stage".

Note that both PCA+LDA and multi-way LDA divide the inter-session compensation into two stages. In the 1st stage, i-vectors are projected into a lower dimensional space via PCA or via the matrix $\mathbf{C}$ in Eq. 7. Then, in the 2nd stage, the dimension of the projected i-vectors is further reduced by LDA to 60.[2] Fig 2 shows the effect of varying the projection dimension in the first stage for both PCA+LDA and multi-way LDA. Evidently, the performance of both methods has a similar trend with respect to this dimension, with multi-way LDA always performs slightly better than PCA+LDA for all projection dimensions. Table 2 also shows that multi-way LDA outperforms PCA+LDA.

7. CONCLUSION

Four techniques aiming to alleviate the small sample-size problem in estimating the LDA and WCCN projection matrices in i-vector based speaker verification have been compared. It was found that utterance partitioning is the most effective way to alleviate the small smaple size problem, followed by multi-way LDA and PCA+LDA.

[2]Because rank $\left(\mathbf{S}_w^{-1}\mathbf{S}_b\right) = \min\{400, S-1\} = 62$, the projected dimension should be set to a value smaller than this rank.

8. REFERENCES

[1] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, May 2011.

[2] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 4, pp. 1435–1447, May 2007.

[3] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308–311, May 2006.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[5] A. Hatch, S. Kajarekar, and A. Stolcke, "Within-class covariance normalization for SVM-based speaker recognition," in *Proc. of the 9th International Conference on Spoken Language Processing*, Pittsburgh, PA, USA, Sep. 2006, pp. 1471–1474.

[6] P. Kenny, "Bayesian speaker verification with heavy-tailed priors," in *Proc. of Odyssey: Speaker and Language Recognition Workshop*, Brno, Czech Republic, June 2010.

[7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[8] W. Zhao, R. Chellappa, and P. Phillips, "Subspace linear discriminant analysis for face recognition," *Technical Report CAR-TR-914*, 1999.

[9] K. Fukunaga, *Introduction to Statistical Pattern Classification*, Academic Press, USA, 1990.

[10] S. Raudys and R. P. W. Duin, "On expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix," *Pattern Recognition Letters*, vol. 19, pp. 385–392, 1998.

[11] M.W. Mak and W. Rao, "Utterance partitioning with acoustic vector resampling for GMM-SVM speaker verification," *Speech Communication*, vol. 53, no. 1, pp. 119–130, Jan. 2011.

[12] W. Rao and M.W. Mak, "Utterance partitioning with acoustic vector resampling for i-vector based speaker verification," in *Proc. of Odyssey: Speaker and Language Recognition Workshop*, Singapore, Jun. 2012.

[13] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, pp. 531–537, 1999.

[14] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. of the Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[15] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, pp. 165–175, 1989.

[16] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *Annals of Statistics*, vol. 23, pp. 73–102, 1995.

[17] L. F. Chen, H. Y. M. Liao, M. T. Ko, J. C. Lin, and G. J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, pp. 1713–1726, 2000.

[18] R. Huang, Q. Liu, H. Lu, and S. Ma, "Solving the small sample size problem of LDA," in *Proc. of International Conference on Pattern Recognition*, 2002, pp. 29–32.

[19] J. Ye, "Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems," *Journal of Machine Learning Research*, vol. 6, no. 1, pp. 483–502, 2005.

[20] J. Ye and T. Xiong, "Computational and theoretical analysis of null space and orthogonal linear discriminant analysis," *The Journal of Machine Learning Research*, vol. 7, pp. 1183–1204, 2006.

[21] N. Dehak, R. Dehak, P. Kenny, N. Brummer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification," in *Proc. Interspeech 2009*, Sep. 2009, pp. 1559–1562.

[22] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 42–54, Jan. 2000.

[23] H.B. Yu and M.W. Mak, "Comparison of voice activity detectors for interview speech in NIST speaker recognition evaluation," in *Proc. of Interspeech 2011*, Florence, Aug. 2011, pp. 2353–2356.

[24] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *J. Acoust. Soc. Am.*, vol. 55, no. 6, pp. 1304–1312, Jun. 1974.

[25] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proc. of Odyssey: Speaker and Language Recognition Workshop*, Crete, Greece, Jun. 2001, pp. 213–218.

[26] M. McLaren and D. van Leeuwen, "Source-normalised LDA for robust speaker recognition using i-vectors from multiple speech sources," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, pp. 755–766, 2012.