INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Locally Weighted Linear Discriminant Analysis for Robust Speaker Verification

Abhinav Misra, Shivesh Ranjan, John H.L. Hansen

Center for Robust Speech Systems (CRSS)
Erik Jonsson School of Engineering & Computer Science
The University of Texas at Dallas (UTD), Richardson, Texas, USA
abhinav.misra@utdallas.edu, shivesh.ranjan@utdallas.edu, john.hansen@utdallas.edu

Abstract

Channel compensation is an integral part for any state-of-the-art speaker recognition system. Typically, Linear Discriminant Analysis (LDA) is used to suppress directions containing channel information. LDA assumes a unimodal Gaussian distribution of the speaker samples to maximize the ratio of the between-speaker variance to within-speaker variance. However, when speaker samples have multi-modal non-Gaussian distributions due to channel or noise distortions, LDA fails to provide optimal performance. In this study, we propose Locally Weighted Linear Discriminant Analysis (LWLDA). LWLDA computes the within-speaker scatter in a pairwise manner and then scales it by an affinity matrix so as to preserve the within-class local structure. This is in contrast to another recently proposed non-parametric discriminant analysis method called NDA. We show that LWLDA not only performs better than NDA but also is computationally much less expensive. Experiments are performed using the DARPA Robust Automatic Transcription of Speech (RATS) corpus. Results indicate that LWLDA consistently outperforms both LDA and NDA on all trial conditions.

Index Terms: Speaker Recognition, Linear Discriminant Analysis, Non-parametric Discriminant Analysis, Speaker Verification, DARPA RATS, Locally Weighted Linear Discriminant Analysis, Nearest-neighbor

1. Introduction

Automatic speaker recognition has made significant advances in recent years [1]. Present state-of-the-art systems employ i-Vectors [2] as the front-end. After i-Vectors have been extracted from an audio signal, several channel compensation methods are employed to reduce the impact of channel distortions. Here, LDA with Fisher criterion [3] is one of the most commonly used channel compensation tools. Post channel compensation based on Probabilistic Linear Discriminant Analysis (PLDA) [4] is used as the back-end to classify the i-Vectors.

LDA aims to compute a reduced set of dimensions onto which i-Vectors can be projected, so that variability between the same-speaker samples can be minimized while at the same time maximizing the variability between different-speaker samples. This is accomplished by maximizing the ratio of the between-speaker covariance to the within-speaker covariance. Sources of within-speaker variation can be different channels, languages, acoustic environments, or speaking styles. These variations cannot only make within-speaker samples distribution multi-modal, but can also introduce a non Gaussian behaviour. In [5], the authors showed that when data in a within-class scatter matrix comes from different channel sources, it is distributed in the form of different clusters, with each cluster corresponding to a separate channel source. In [4], the author clearly showed how a Gaussian assumption on the distribution of i-Vectors leads to sub-optimal performance in speaker recognition systems.

Recently, NDA was proposed by authors in [6], as a tool to suppress above noted isuues. However, in this study, we observed that NDA does not always provide improvement over LDA. Furthermore, NDA is computationally very expensive since it involves computing nearest neighbors of a speaker i-Vector in each of the other speaker classes. Motivated by these observations, we present an alternative non-parametric discriminant analysis technique (LWLDA) that measures the within-speaker variation on a local basis using an affinity matrix. The affinity matrix is chosen such that nearby data pairs in the within-speaker scatter are kept closer, while the far apart data pairs are not imposed to be close to each other. Weighing the within-speaker data in such a way helps in locally preserving its multimodal property, and improves the system performance. Since, we do not need affinity values for speaker i-Vectors belonging to different classes, this highly contributes to reducing the computational costs. Furthermore, because of the pairwise non-parametric computation, the between-class scatter is generally full rank with no eigenvalue multiplicity. Hence, the proposed LWLDA can be employed for dimensionality reduction into any dimensional spaces.

We evaluate LWLDA against LDA and NDA using DARPA RATS corpora. The RATS data consists of conversational telephone recordings that are retransmitted over a Multi Radio-Link Channel System (MRLC) containing 8 channels [7]. Each of these channels have different degrees and types of distortion characteristics. We consider the following enrollment-test conditions of RATS data: 120s, 30s and 10s. We show improvement using LWLDA on all of these conditions. It can be noted that we have shown improvement with LWLDA over LDA and NDA on NIST Speaker Recognition Evaluation (SRE) data as well in [8].

The remainder of the paper is as follows. We first give a brief review of how LDA and NDA are used in speaker verification systems. Section III presents the proposed LWLDA method. Section IV presents the experimental setup and Section V discusses results of this study. Section VI concludes the paper and discusses future work.

This project was funded by AFRL under contract FA8750-15-1-0205 and partially by the University of Texas at Dallas from the Distinguished University Chair in Telecommunications Engineering held by J. H. L. Hansen.

http://dx.doi.org/10.21437/Interspeech.2017-581

2. Parametric and Non-parametric Discriminant Analysis

2.1. LDA

LDA attempts to maximize the discrimination between different speaker i-Vectors by finding a set of dimensions where between-speaker covariance is maximum while while at the same time minimizing the within-speaker covariance. This set of dimensions is obtained with the following procedure: First, the between-speaker and within-speaker covariance matrices, S_b and S_w respectively, are computed as,

$$\boldsymbol{S_b} = \frac{1}{n}\sum_{spk=1}^{p} n_{spk}(\boldsymbol{\mu_{spk}} - \boldsymbol{\mu})(\boldsymbol{\mu_{spk}} - \boldsymbol{\mu})^t. \quad (1)$$

$$\boldsymbol{S_w} = \frac{1}{n}\sum_{spk=1}^{p}\sum_{j=1}^{n_{spk}} (\boldsymbol{\omega_j^{spk}} - \boldsymbol{\mu_{spk}})(\boldsymbol{\omega_j^{spk}} - \boldsymbol{\mu_{spk}})^t. \quad (2)$$

where, the number of speakers (or classes) is p. $\boldsymbol{\omega}$ is an i-Vector and n_{spk} is the number of i-Vectors corresponding to a speaker spk. $\boldsymbol{\mu_{spk}}$ is the mean of i-Vectors belonging to speaker spk, while $\boldsymbol{\mu}$ is the global mean of all the n i-Vectors present in the development data-set.

In order to formulate a criterion for class separability, after computation of the scatter matrices, we need to convert them to a number. This number should be larger when the between-class scatter is larger or the within-class scatter is smaller. One typical criteria is:

$$f = \mathrm{tr}(\boldsymbol{S_w}^{-1}\boldsymbol{S_b}). \quad (3)$$

Our aim here is to optimize f by finding a linear transformation $\boldsymbol{A}$, such that $\boldsymbol{A}$ would transform the i-Vectors from $\boldsymbol{x}$ dimensions to $\boldsymbol{y}$ dimensions ($\boldsymbol{y} < \boldsymbol{x}$) as:

$$\boldsymbol{y} = \boldsymbol{A}^t\boldsymbol{x}. \quad (4)$$

It turns out that the value of $\boldsymbol{A}$ that optimizes j, is given by the eigenvectors corresponding to the largest eigenvalues of $\boldsymbol{S_w}^{-1}\boldsymbol{S_b}$ [3].

2.2. NDA

NDA accomplishes the same objective as LDA, except now, instead of a global average, we consider a local sample mean. The Local mean for an individual sample of a class is computed by averaging the sample's k nearest neighbors in other class [9],

$$\boldsymbol{S_b} = \sum_{spk=1}^{p}\sum_{\substack{l=1,\\ l\neq spk}}^{p}\sum_{j=1}^{n_{spk}} w_j^{spk,l}(\boldsymbol{\omega_j^{spk}} - \boldsymbol{M_j^{spk,l}})(\boldsymbol{\omega_j^{spk}} - \boldsymbol{M_j^{spk,l}})^t. \quad (5)$$

where, $w_j^{spk,l}$ is the weighting function, and $\boldsymbol{M_j^{spk,l}}$ is the local mean of k-NN samples for $\boldsymbol{\omega_j^{spk}}$ from class l, given as:

$$\boldsymbol{M_j^{spk,l}} = \frac{1}{k}\sum_{q=1}^{k} NN_q(\boldsymbol{\omega_j^{spk}}, l). \quad (6)$$

where, $NN_q(\boldsymbol{\omega_j^{spk}}, l)$ is the q^{th} nearest neighbor of $\boldsymbol{\omega_j^{spk}}$ in class l.

S_w is computed in a similar way as S_b, except the weighting function is set to 1 and local gradients are computed within each class. Once we obtain the scatter matrices, just like LDA, $\boldsymbol{A}$ is computed as the eigenvectors corresponding to largest eigenvalues of $\boldsymbol{S_w}^{-1}\boldsymbol{S_b}$.

3. Locally Weighted Linear Discriminant Analysis

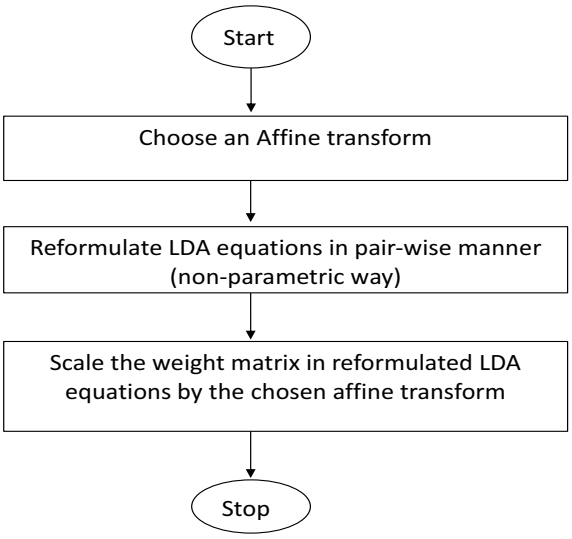

Figure 1: *Steps involved in LWLDA computation.*

In this section, we present a new method entitled localized weighted linear discriminant analysis (LWLDA). LWLDA, like NDA, is based on non-parametric discriminant analysis. However, unlike NDA, it focusses on weighing the within-speaker i-Vectors. The weight matrix is computed such that the complex (multi-modal) structure of the within-speaker data is preserved. This is achieved by constraining the values of weight matrix to be between 0 and 1. The values are large if i-Vectors are close, while small if i-Vectors are far apart. Hence, far apart sample pairs belonging to the same class will have less influence on the within-speaker scatter computation as compared to closer sample pairs. Sample pairs belonging to different classes are not weighted by the weight/affinity matrix. This occurs since we want them to be separated from each other, irrespective of whether any affinity exists between them or not. Fig 1 shows the basic steps involved in LWLDA computation.

3.1. Choice of Affinity Matrix

One of the easiest choices of an affinity matrix $\boldsymbol{H}$ can be: assign $H_{i,j} = 1$, when i-Vectors are neighbors and $H_{i,j} = 0$, otherwise. However, this kind of hard thresholding does not represent the contribution that far apart i-Vectors might have in S_w computation. Hence, we consider a Gaussian function that varies with the local density h of data samples, as our affinity matrix.

$$H_{i,j} = \exp(-\frac{||\boldsymbol{\omega_i} - \boldsymbol{\omega_j}||^2}{h_i h_j}), \quad (7)$$

where, h_i represents a scaling factor that takes into account the distribution of samples around $\boldsymbol{\omega_i}$. It is defined as:

$$h_i = ||\boldsymbol{\omega_i} - \boldsymbol{\omega_i^k}||. \quad (8)$$

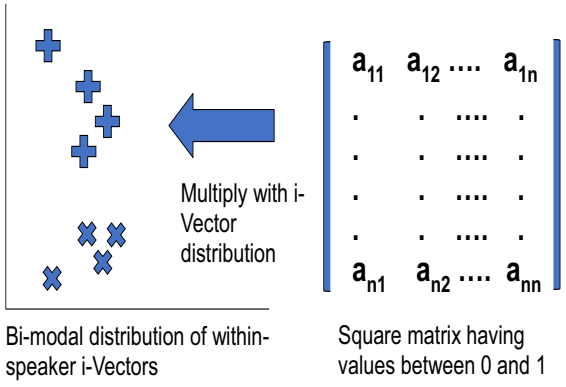

Figure 2: *LWLDA: An example showing affinity matrix being multiplied with a within-speaker scatter that has bi-modal distribution. Here, n = number of within-speaker i-Vectors.*

where, $\boldsymbol{\omega}_i^k$ is the k-th nearest neighbour of $\boldsymbol{\omega_i}$. The value of k is derived heuristically, and can vary for different distributions. Fig 2 shows a synthetic example that illustrates the process of an affinity matrix being multiplied with a bi-modal within-scatter distribution.

3.2. Lemma

Once an affinity matrix is chosen, next we need to incorporate it in LDA equations. To accomplish that, we reformulate LDA equations in a nonparametric manner, using data pairs.

The new $\boldsymbol{S_b}$ and $\boldsymbol{S_w}$ are given by:

Lemma:

$$\boldsymbol{S'_b} = \frac{1}{2}\sum_{i,j=1}^{n} \boldsymbol{W^b_{i,j}}(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})^t. \tag{9}$$

$$\boldsymbol{S'_w} = \frac{1}{2}\sum_{i,j=1}^{n} \boldsymbol{W^w_{i,j}}(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})^t. \tag{10}$$

where,

$$\boldsymbol{W^w_{i,j}} = \begin{cases} \frac{1}{n_{spk}} & z_i = z_j = spk \\ 0 & z_i \neq z_j. \end{cases} \tag{11}$$

$$\boldsymbol{W^b_{i,j}} = \begin{cases} \frac{1}{n} - \frac{1}{n_{spk}} & z_i = z_j = spk \\ \frac{1}{n} & z_i \neq z_j. \end{cases} \tag{12}$$

where, z are the speaker class labels.

It can be observed from the new formulation that $(\frac{1}{n} - \frac{1}{n_{spk}})$ is negative, while $\frac{1}{n}$ and $\frac{1}{n_{spk}}$ are positive. Hence, when data pairs belong to the same class, the terms in $\boldsymbol{S'_b}$ are weighed negatively making $\boldsymbol{S'_b}$ smaller, while terms in $\boldsymbol{S'_w}$ are weighed positively making $\boldsymbol{S'_w}$ larger. The exact opposite happens for the case where data pairs belong to different classes. Terms in $\boldsymbol{S'_b}$ are weighed positively making $\boldsymbol{S'_b}$ larger, while terms in $\boldsymbol{S'_w}$ are given zero weight making $\boldsymbol{S'_w}$ smaller. Therefore, the new formulation conforms to our notion of LDA, where the distance between samples of different classes is maximized, while distance between samples of the same classes are minimized.

The transformation matrix $\boldsymbol{A}$ is computed the same way as in the original LDA formulation, except now the local $\boldsymbol{S'_b}$ and $\boldsymbol{S'_w}$ are used instead of the global $\boldsymbol{S_b}$ and $\boldsymbol{S_w}$ respectively. As a result of the affinity transform, $\boldsymbol{S'_b}$ generally has a much higher rank than $p-1$. If the affinity value is set to 1 for all sample pairs, $\boldsymbol{S'_b}$ and $\boldsymbol{S'_w}$ become equal to $\boldsymbol{S_b}$ and $\boldsymbol{S_w}$ respectively. Hence, we can say that LWLDA is effectively a localized variant of LDA.

In this study, we prove the Lemma for the specific case of our choice of our weight matrix. A more general proof is given in [10].

3.3. Proof of Lemma

Rewriting Eq. (2) as:

$$\begin{aligned} \boldsymbol{S_w} &= \sum_{spk=1}^{p}\sum_{i=1}^{n_{spk}}(\boldsymbol{\omega_i} - \frac{1}{n_{spk}}\sum_{j=1}^{n_{spk}}\boldsymbol{\omega_j})(\boldsymbol{\omega_i} - \frac{1}{n_{spk}}\sum_{j=1}^{n_{spk}}\boldsymbol{\omega_j})^t \\ &= \sum_{i=1}^{n}\boldsymbol{\omega_i}\boldsymbol{\omega_i}^t - \sum_{spk=1}^{p}\frac{1}{n_{spk}}\sum_{i=j}^{n_{spk}}\boldsymbol{\omega_i}\boldsymbol{\omega_j}^t\sum_{i\neq j}^{n_{spk}}\boldsymbol{\omega_i}\boldsymbol{\omega_j}^t \\ &- \sum_{spk=1}^{p}\frac{1}{n_{spk}}\sum_{i=j}^{n_{spk}}\boldsymbol{\omega_j}\boldsymbol{\omega_i}^t\sum_{i\neq j}^{n_{spk}}\boldsymbol{\omega_j}\boldsymbol{\omega_i}^t + \sum_{spk=1}^{p}\frac{1}{n_{spk}^2}\sum_{j=1}^{spk}\boldsymbol{\omega_j}\boldsymbol{\omega_j^t}. \end{aligned} \tag{13}$$

Let us assume $\frac{1}{n_{spk}} = \boldsymbol{W_{i,j}}$. Also, from our choice of affinity matrix we know that the diagonal elements of $\boldsymbol{W_{i,j}}$ will be ones. Hence, the above equation can be written as:

$$\begin{aligned} \boldsymbol{S_w} &= \sum_{i=1}^{n}\boldsymbol{W^w_{i,i}}\boldsymbol{\omega_i}\boldsymbol{\omega_i}^t - \sum_{i,j=1}^{n}\boldsymbol{W^w_{i,j}}\boldsymbol{\omega_i}\boldsymbol{\omega_j}^t \\ &- \sum_{i,j=1}^{n}\boldsymbol{W^w_{i,j}}\boldsymbol{\omega_j}\boldsymbol{\omega_i}^t + \sum_{j=1}^{n}(\boldsymbol{W^w_{j,j}})^2\boldsymbol{\omega_j}\boldsymbol{\omega j}^t \\ &= \sum_{i=1}^{n}\boldsymbol{W^w_{i,i}}\boldsymbol{\omega_i}\boldsymbol{\omega_i}^t - \sum_{i,j=1}^{n}\boldsymbol{W^w_{i,j}}\boldsymbol{\omega_i}\boldsymbol{\omega_j}^t \\ &- \sum_{i,j=1}^{n}\boldsymbol{W^w_{i,j}}\boldsymbol{\omega_j}\boldsymbol{\omega_i}^t + \sum_{j=1}^{n}\boldsymbol{W^w_{j,j}}\boldsymbol{\omega_j}\boldsymbol{\omega j}^t \\ &= \sum_{i,j=1}^{n}\boldsymbol{W^w_{i,j}}(\boldsymbol{\omega_i}\boldsymbol{\omega_i}^t + \boldsymbol{\omega_j}\boldsymbol{\omega_j}^t - \boldsymbol{\omega_i}\boldsymbol{\omega j}^t - \boldsymbol{\omega_j}\boldsymbol{\omega i}^t) \\ &= \sum_{i,j=1}^{n}\boldsymbol{W^w_{i,j}}(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})^t. \end{aligned} \tag{14}$$

Next, we know that the scatter matrix for the entire distribution is the sum of the between-class and within-class scatter matrices:

$$\boldsymbol{S_m} = \boldsymbol{S_b} + \boldsymbol{S_w} = \sum_{i=1}^{n}(\boldsymbol{\omega_i} - \boldsymbol{\mu})(\boldsymbol{\omega_i} - \boldsymbol{\mu})^t. \tag{15}$$

Hence, we have:

$$\begin{aligned} \boldsymbol{S_b} &= \sum_{i=1}^{n}(\boldsymbol{\omega_i} - \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{\omega_j})(\boldsymbol{\omega_i} - \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{\omega_j})^t - \boldsymbol{S_w} \\ &= \sum_{i,j=1}^{n}(\frac{1}{n} - \boldsymbol{W^w_{i,j}})(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})(\boldsymbol{\omega_i} - \boldsymbol{\omega_j})^t. \end{aligned} \tag{16}$$

3.4. Application of Affinity Matrix

Finally, after reformulating the LDA equations in a pairwise manner, we apply affinity transform $\boldsymbol{H}$. This is accomplished by simply replacing $\boldsymbol{W}$ by $\boldsymbol{H}$ in the above equations.

Table 1: *Speaker recognition performance in terms of all the DARPA metrics (all values in %).*

Eval. Cond.	Channel Compensation Methods	EER	Miss @4%FA	Miss @2.5%FA	Miss @5%FA	Miss @1.5%FA	Miss @1%FA	Miss @3%FA	FA @10%Miss	FA @3%Miss
120s-120s	LDA	5.32	6.23	8.14	5.47	10.55	12.81	7.18	1.67	12.96
	NDA	5.51	6.73	8.81	5.84	11.68	14.45	7.96	2.00	14.66
	LWLDA	5.04	5.84	7.87	5.08	10.63	13.48	7.02	1.68	12.33
30s-30s	LDA	8.16	13.48	17.56	11.76	22.41	26.94	15.80	6.40	21.65
	NDA	8.52	14.53	19.13	12.69	24.72	29.15	17.28	6.96	23.00
	LWLDA	7.54	12.24	16.53	10.40	22.04	26.62	14.95	5.22	18.58
10s-10s	LDA	18.46	44.65	52.78	40.86	61.02	66.87	49.40	32.18	60.56
	NDA	19.69	47.18	55.42	43.23	63.23	68.39	52.06	35.79	64.14
	LWLDA	17.99	42.99	50.55	39.29	59.01	65.01	47.54	30.98	58.56

4. Experiments

We conduct our experiments using the corpora available as part of RATS speaker recognition task. The data distributed by Linguistic Data Consortium (LDC) in the form of: LDC2012E49, LDC2012E63, LDC2012E69, LDC2012E85, LDC2012E117, contain speech in five languages: Levantine Arabic, Dari, Farsi, Pashto, and Urdu. We divide these data into three parts for our system training, enrolment and test. There are a total of 305 speakers available for evaluation (enrolment and test), while 5913 speakers are set aside for training system hyperparameters including Universal Background Model (UBM), Total Variability (T.V.) Matrix and PLDA. All speakers represent both male and female genders. There are 8 channels (A-H) through which each of the speaker's telephone recordings are retransmitted. In addition to 8 channels, there is also the speaker's original telephone channel recording. Our evaluation and training data-sets contain recordings from all 9 channels. Each speaker model is trained using all sessions coming from 8 extremely degraded communication channels as well as the original telephone channel recording. A trial is designed using one speaker model and one test session. The test sessions are also chosen to represent all 9 sources of speech recordings. To evaluate system performance, we consider 3 duration-specific tasks with the following enrollment-test conditions: 120s, 30s, 10s. The total number of trials for each condition are created by taking a cartesian product of all the enrolment and test sessions. This leads to roughly 3.2 million trials, with 10,617 target trials and 3,227,568 non-target trials. It can be noted here, that in all the enrollment-test conditions, system hyperparameters are all trained using complete recordings. DARPA carried out the RATS project in five phases using a list of metrics to measure the system performance. These metrics, apart from measuring the Equal Error Rate (EER), also measure the Miss probability and False Alarm (FA) at different points on the Detection Error Trade-off (DET) curve. In this study, we report our results using all the performance metrics as identified by DARPA in the RATS project.

The speech is parameterized using 60-dimensional Mel-Frequency Cepstral Coefficients (MFCCs) containing delta and delta-delta coefficients. A gender independent, 2048 component, full covariance UBM is trained using 55,982 recordings representing 5913 speakers. Each of these recordings are roughly 15 min. long. The zero and first order statistics are computed for each of the recordings from UBM, and then used to compute a 600 dimensional T.V. matrix. Next, we extract 600-dimensional i-Vectors from the T.V. matrix and apply LDA, NDA or LWLDA for channel compensation. After channel compensation, we perform length normalization [11] and finally classify the i-Vectors using PLDA scoring.

5. Results

Table 1 shows performance of our system using all three channel compensation methods. All results are obtained by using the value of k, that gives optimal performance. It can be observed that LWLDA gives better performance than LDA and NDA in all enrollment-test conditions. In terms of EER, LWLDA provides 7.6% relative improvement over LDA and 11.50% relative improvement over NDA, in case of 30s-30s evaluation condition. In terms of FA@10%Miss, the improvement increases upto 18.44% over LDA and 25.0% over NDA. Other evaluation conditions also show improvement with LWLDA, although it is smaller than that obtained in case of 30s-30s evaluation condition.

5.1. Computational Complexity

NDA computation has 4 nested loops as can be observed from Eq. 5. This leads to a computational time complexity of $O(p(p-1)n_{spk}k)$, which simplifies to $O(p^2 n_{spk}k)$. On the other hand, LWLDA has only two nested loops with a time complexity of $O(pn_{spk})$. Hence, we can observe that NDA involves much more computations than LWLDA.

6. Conclusion

In this study, we have considered the problem of multi-modality for within-speaker i-Vectors that is caused due to channel or noise distortions. Even though LDA has become an integral part of many state-of-the-art speaker recognition systems, it still fails to cope with the multi-modality issue. Recently, NDA was proposed to address this problem, however, we observed in our study that NDA falls short of providing sufficient gains in system performance. Motivated by these observations, we proposed an alternative way of computing the scatter matrices. Using our proposed method of LWLDA, we obtained consistent gains on different speaker recognition tasks employing DARPA RATS data. We showed that not only is LWLDA better than NDA, it is also much less computationally expensive. The fact that LWLDA leads to dimensionality reduction upto any dimensional spaces, can be extremely useful in areas like language recognition, where the number of classes (languages) is usually less than the number of i-Vector dimensions. Recently, NIST conducted the SRE-2016 evaluation. Due to multi-modality of different data-sets being used to develop systems, LWLDA might provide gains to the systems used for SRE-2016 evaluations. Hence, in future, we intend to apply LWLDA on SRE-2016 data as well.

7. References

[1] J. H. L. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial review," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74–99, Nov 2015.

[2] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio Speech Lang. Process.*, May 2010.

[3] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. 2nd ed. New York: Academic Press, 1990, ch.10.

[4] P. Kenny, "Bayesian speaker verification with heavy-tailed priors," in *Proc. Odyssey*, 2010.

[5] M. McLaren and D. van Leeuwen, "Source-normalized lda for robust speaker recognition using i-vectors from multiple speech sources," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 755–766, March 2012.

[6] S. Sadjadi, J. Pelecanos, and W. Zhu, "Nearest neighbor discriminant analysis for robust speaker recognition," in *Proc. Interspeech*, 2014.

[7] K. Walker and S. Strassel, "The rats radio traffic collection system," in *Proc. Odyssey*, 2012.

[8] A. Misra and J. H. Hansen, "Compensating for language mismatch in speaker verification," *Speech Communication*, submitted and under review.

[9] K. Fukunaga and J. Mantock, "Nonparametric discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 6, pp. 671–678, 1983.

[10] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007. [Online]. Available: http://dl.acm.org/citation.cfm?id=1248659.1248694

[11] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-Vector length normalization in speaker recognition systems," in *Proc. InterSpeech*, Florence, Italy, Oct. 2011.