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

Domain mismatch modeling of out-domain i-vectors for PLDA speaker verification

Md Hafizur Rahman , Ivan Himawan , David Dean , Sridha Sridharan

Speech and Audio Research Laboratory, Queensland University of Technology, Brisbane, Australia
{m20.rahman, i.himawan, s.sridharan}@qut.edu.au, ddean@ieee.org

Abstract

The state-of-the-art i-vector based probabilistic linear discriminant analysis (PLDA) trained on non-target (or out-domain) data significantly affects the speaker verification performance due to the domain mismatch between training and evaluation data. To improve the speaker verification performance, sufficient amount of domain mismatch compensated out-domain data must be used to train the PLDA models successfully. In this paper, we propose a domain mismatch modeling (DMM) technique using maximum-a-posteriori (MAP) estimation to model and compensate the domain variability from the out-domain training i-vectors. From our experimental results, we found that the DMM technique can achieve at least a 24% improvement in EER over an out-domain only baseline when speaker labels are available. Further improvement of 3% is obtained when combining DMM with domain-invariant covariance normalization (DICN) approach. The DMM/DICN combined technique is shown to perform better than in-domain PLDA system with only 200 labeled speakers or 2,000 unlabeled i-vectors.

Index Terms: Speaker verification, domain adaptation, GPLDA, DMM, DICN

1. Introduction

In the past few years, speaker verification system has been advanced significantly especially after the introduction of i-vector based PLDA system by Kenny [1]. However, recent studies have found that the dataset mismatch could significantly affect the speaker verification performance. This mismatch happens when speaker verification trained on a particular domain is evaluated on a different domain. The performance gap due to domain mismatch was first introduced as a challenge in the Summer Workshop of 2013 at Johns Hopkins University (JHU) [2]. Results presented in that workshop outlined the discrepancy in performance caused by domain variability and this problem is later adopted as domain mismatch by the research community

Recently, researchers have introduced various domain-mismatch compensation techniques in order to achieve the state-of-the-art speaker verification performance with limited adaptation data. Based on the availability of speaker labels these techniques can be classified into supervised and unsupervised techniques. For supervised domain adaptation, Garcia-Romero *et al.* [3] proposed four similarly well performing PLDA parameter adaptation techniques called fully Bayesian adaptation, approximate MAP adaptation, weighted likelihood and parameter interpolation method. They concluded that training the universal background model (UBM) and total variability on out-domain data has negligible effect on the overall system performance. In order to transfer the domain information from target to training domain, Wang *et al.* [4] proposed two maximum likelihood linear transformation (MLLT) approaches for adapting PLDA parameters. Hong *et al.* [5] proposed another transfer learning approach based on Bayesian joint probability, where they used Kullback-Leibler (KL) divergence to maximize the optimization function for PLDA parameters adaptation. These transfer learning based methods performed better than the interpolated PLDA parameter adaptation by Garcia-Romero [3]. Villalba *et al.* [6] investigated a supervised fully Bayesian approach and variational approximation to compute the intractable posterior using conjugate priors for PLDA parameter adaptation. They found a very good performance due to the adaptation of the channel matrix. In order to compensate the mismatch in the i-vector subspace, Aronowitz [7] introduced inter dataset variability compensation (IDVC) to shift the dataset in the i-vector subspace based on nuisance attribute projection (NAP) and found relatively better performance compared to PLDA parameter adaptation. Singer *et al.* [8] proposed a library whitening technique for dataset variability compensation trained on out-domain data. This approach automatically adjusts the whitening scheme to compensate the domain variation from the out-domain data.

The unsupervised domain adaptation is a more challenging task due to the unavailability of the speaker labels. For unsupervised adaptation, Garcia-Romero *et al.* [9] introduced agglomerative hierarchical clustering (AHC) method for clustering the unlabeled data. They successfully retrieved 23% of the original labels and used those clustered data for PLDA parameter adaptation. This approach lacks in performance due to the training of PLDA parameters with small amount retrieved in-domain labels. In their other work, Garcia-Romero *et al.* [10] utilized this AHC clustering method to explore the DNN/i-vector based system. They used a DNN to collect the sufficient statistics for i-vector extraction and achieved remarkable unsupervised adaptation performance. Shum *et al.* [11] also investigated different unsupervised clustering methods including AHC, Infomap and Markov Clustering (MCL), and used interpolated PLDA parameter adaptation similar to [3]. Villalba *et al.* [12] trained a generative model using unknown labels modeled as latent variables and used the variational Bayes approach to predict the posterior distributions of the latent variables. Glembek *et al.* [13] proposed a within-speaker covariance correction (WCC) approach for adaptation of the LDA subspace, which has proven to be very effective for unsupervised adaptation. In our recent work, we proposed dataset-invariant covariance normalization (DICN) [14] approach to capturing domain mismatch from the global mean i-vector and later compensated this variability from the PLDA training i-vectors.

In this paper, we propose domain mismatch modeling (DMM) technique to model the domain mismatch from out-domain i-vectors. This work is motivated by the i-vector denoising approach proposed by Kheder *et al.* [15], where they estimated the clean i-vectors directly from their noisy version using MAP estimation. However, in this work we first model

  http://dx.doi.org/10.21437/Interspeech.2017-668

the domain mismatch from all out-domain i-vectors using MAP estimation and later use the expectation maximization (EM) to update the models more accurately. The new mismatch compensated out-domain i-vectors are then estimated by the difference between original i-vectors and domain mismatch. We also combine the DICN approach with DMM to compensate the domain mismatch further in the i-vector subspace to train the PLDA models more efficiently.

The rest of this paper is organized as follows: Section 2 details the DMM approach. Section 3 describes the DICN approach and Section 4 explains the length-normalized GPLDA modeling. The experimental protocol and corresponding results are described in Section 5 and Section 6. Finally, Section 7 concludes the paper.

2. Domain mismatch modeling (DMM)

In i-vector paradigm [16], the speaker and session dependent GMM super-vector $\mathbf{M}$ is represented as follows,

$$\mathbf{M} = \mathbf{m} + \mathbf{T}\mathbf{w}, \tag{1}$$

where $\mathbf{m}$ is the mean super-vector of the UBM, $\mathbf{T}$ is a low rank total-variability matrix and $\mathbf{w}$ is the total variability factor assumed to be normally distributed $\mathcal{N}(0, \mathbf{I})$. The training of the $\mathbf{T}$ matrix is similar to the eigenvoice modeling in joint factor analysis (JFA), except in this case each recording is assumed to be coming from different speakers.

In DMM approach, the domain mismatch is first modeled using MAP estimation and later compensated from the out-domain i-vectors to get the adapted i-vectors. For a given out-domain i-vector $\mathbf{w}_{old}$, the domain mismatch compensated i-vector $\mathbf{w}_{new}$ is estimated by,

$$\mathbf{w}_{new} = \mathbf{w}_{old} - \mathbf{w}_d \tag{2}$$

where $\mathbf{w}_d$ is the domain mismatch part of the original out-domain i-vector. Both $\mathbf{w}_d$ and $\mathbf{w}_{new}$ are assumed to be normally distributed and their probability density functions are defined as,

$$p(\mathbf{w}_d) = \mathcal{N}(\mathbf{m}_d, \mathbf{\Sigma}_d) \tag{3}$$

$$p(\mathbf{w}_{new}) = \mathcal{N}(\mathbf{m}_{new}, \mathbf{\Sigma}_{new}) \tag{4}$$

Now, given out-domain i-vector $\mathbf{w}_{old}$, the posterior distribution of $\mathbf{w}_d$ is defined as follows,

$$\begin{aligned} p(\mathbf{w}_d|\mathbf{w}_{old}) &= \frac{p(\mathbf{w}_{old}|\mathbf{w}_d)\, p(\mathbf{w}_d)}{p(\mathbf{w}_{old})} && (5) \\ &\propto p(\mathbf{w}_{old}|\mathbf{w}_d)\, p(\mathbf{w}_d) && (6) \end{aligned}$$

Using a MAP estimation, we can now model this domain mismatch $\hat{\mathbf{w}}_d$ as follows,

$$\begin{aligned} \hat{\mathbf{w}}_d &= \underset{\mathbf{w}_d}{\operatorname{argmax}} \ln p(\mathbf{w}_d|\mathbf{w}_{old}) && (7) \\ &= \underset{\mathbf{w}_d}{\operatorname{argmax}} \{\ln p(\mathbf{w}_{old}|\mathbf{w}_d) + \ln p(\mathbf{w}_d)\} && (8) \end{aligned}$$

The probability density function $p(\mathbf{w}_{old}|\mathbf{w}_d)$ can be presented as follows,

$$\begin{aligned} p(\mathbf{w}_{old}|\mathbf{w}_d) = (2\pi)^{D/2} \times |\mathbf{\Sigma}_i|^{-1/2} \\ \times e^{-1/2 \times (\mathbf{w}_{old} - \mathbf{w}_d - \mathbf{m}_i)^T \mathbf{\Sigma}_i^{-1} \times (\mathbf{w}_{old} - \mathbf{w}_d - \mathbf{m}_i)} \end{aligned} \tag{9}$$

where $\mathbf{m}_i$ is the mean and $\mathbf{\Sigma}_i$ is the covariance matrix of the in-domain data; D is the dimension of the i-vectors.

Finally, the domain mismatch $\hat{\mathbf{w}}_d$ is determined by solving the following equation,

Algorithm 1: EM Algorithm for DMM training

Input : $\mathbf{w}_{in} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_M\}$
$\mathbf{w}_{old} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N\}$
Output: $\mathbf{w}_{new} = \mathbf{w}_{old} - \mathbf{w}_d$
Initialization: Estimate $\mathbf{m}_i$ and $\mathbf{\Sigma}_i$ from $\mathbf{w}_{in}$
$\mathbf{m}_d = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{w}_{old}^j - \mathbf{m}_i)$
$\mathbf{\Sigma}_d = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{w}_{old}^j - \mathbf{m}_i)(\mathbf{w}_{old}^j - \mathbf{m}_i)^T$
begin
E-Step:
Compute
$\mathbf{w}_d = (\mathbf{\Sigma}_d^{-1} + \mathbf{\Sigma}_i^{-1})^{-1}[\mathbf{\Sigma}_i^{-1}(\mathbf{w}_{old} - \mathbf{m}_i) + \mathbf{\Sigma}_d^{-1}\mathbf{m}_d]$
M-Step:
$\mathbf{m}_d = \frac{1}{N}\sum_{j=1}^{N}\mathbf{w}_d^j$
$\mathbf{\Sigma}_d = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{w}_d^j - \mathbf{m}_d)(\mathbf{w}_d^j - \mathbf{m}_d)^T$
until convergence

$$\frac{\partial}{\partial \mathbf{w}_d}\{\ln p(\mathbf{w}_{old}|\mathbf{w}_d) + \ln p(\mathbf{w}_d)\} = 0 \tag{10}$$

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}_d}\{(\mathbf{w}_{old} - \mathbf{w}_d - \mathbf{m}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{w}_{old} - \mathbf{w}_d - \mathbf{m}_i) \\ + (\mathbf{w}_d - \mathbf{m}_d)^T \mathbf{\Sigma}_d^{-1}(\mathbf{w}_d - \mathbf{m}_d)\} = 0 \end{aligned} \tag{11}$$

After the derivation it becomes,

$$-\mathbf{\Sigma_i}^{-1}(\mathbf{w}_{old} - \hat{\mathbf{w}}_d - \mathbf{m}_i) + \mathbf{\Sigma}_d^{-1}(\hat{\mathbf{w}}_d - \mathbf{m}_d) = 0$$

$$\hat{\mathbf{w}}_d = (\mathbf{\Sigma}_d^{-1} + \mathbf{\Sigma}_i^{-1})^{-1}[\mathbf{\Sigma}_i^{-1}(\mathbf{w}_{old} - \mathbf{m}_i) + \mathbf{\Sigma}_d^{-1}\mathbf{m}_d] \tag{12}$$

In the beginning it is difficult to calculate the mean $\mathbf{m}_d$ and covariance $\mathbf{\Sigma}_d$ of the domain mismatch, so we calculated them initially as follows,

$$\mathbf{m}_d = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{w}_{old}^j - \mathbf{m}_i) \tag{13}$$

$$\mathbf{\Sigma}_d = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{w}_{old}^j - \mathbf{m}_i)(\mathbf{w}_{old}^j - \mathbf{m}_i)^T \tag{14}$$

where N is the number of out-domain i-vectors.

Later, an expectation-maximization (EM) algorithm is used to estimate the $\mathbf{m}_d$ and $\mathbf{\Sigma}_d$ more accurately as described in Algorithm 1.

3. DICN approach

After the domain mismatch compensation from the out-domain i-vectors as described in Section 2, we applied DICN [14] transformation to compensate the domain mismatch more efficiently in the i-vector subspace. In this approach, the mismatch is captured and modeled from the global mean i-vector as follows,

$$\mathbf{\Sigma}_{DICN} = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{w}_m - \bar{\mathbf{w}})(\mathbf{w}_m - \bar{\mathbf{w}})^T \tag{15}$$

where M is the total number of i-vectors (in-domain and out-domain) and $\bar{\mathbf{w}}$ is the global mean, which can be calculated as follows,

$$\bar{\mathbf{w}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{w}_m \tag{16}$$

The DICN decorrelated matrix $\mathbf{D}$ is calculated using the Cholesky decomposition of $\mathbf{D}\mathbf{D}^T = \mathbf{\Sigma}_{DICN}^{-1}$. After the estimation of the projection matrix $\mathbf{D}$, the DICN compensated out-domain i-vectors are extracted as follows,

$$\mathbf{w}_{DICN} = \mathbf{D}^T\mathbf{w}_{out} \tag{17}$$

After the estimation of DICN compensated i-vectors, the LDA projection is applied to compensate the session variation and to reduce the dimension of the i-vectors before GPLDA modeling.

4. Length-normalized GPLDA

Instead of compensating the session variability in the i-vector subspace, it more convenient to model the speaker and session variability in the PLDA subspace [1]. The PLDA modeling was first introduced in speaker verification by Kenny [1]. Later, Garcia-Romero [17] introduced GPLDA using length normalized i-vectors, which is computationally more efficient than heavy-tailed assumption [1]. In GPLDA modeling, the length normalized i-vectors can be presented as follows,

$$\mathbf{w}_q = \bar{\mathbf{w}} + \mathbf{U}_1\mathbf{x}_1 + \boldsymbol{\epsilon}_q, \tag{18}$$

where for any given speaker recordings $q = 1, ...Q$; $\mathbf{U}_1$ is the eigenvoice matrix; $\bar{\mathbf{w}} + \mathbf{U}_1\mathbf{x}_1$ is the speaker part with covariance matrix of $\mathbf{U}_1\mathbf{U}_1^T$ and $\boldsymbol{\epsilon}_q$ is the channel part with covariance matrix of $\boldsymbol{\Lambda}^{-1}$, where $\boldsymbol{\Lambda}$ is the precision matrix with full rank.

The scoring between the target and test i-vectors is calculated using the batch likelihood ratio [1]. For given target i-vector $\mathbf{w}_{target}$ and test i-vector $\mathbf{w}_{test}$, the batch likelihood ratio can be calculated as follows,

$$\ln \frac{P(\mathbf{w}_{target}, \mathbf{w}_{test} \mid H_1)}{P(\mathbf{w}_{target} \mid H_0)P(\mathbf{w}_{test} \mid H_0)} \tag{19}$$

where H_1: The speakers are same, H_0: The speakers are different

5. Experimental Setup

In this paper, we evaluated our proposed methods on NIST 2008 short2-short3 *telephone-telephone* and NIST 2010 core-core *telephone-telephone* conditions. The performances were measured using equal error rate (EER) and minimum decision cost function (DCF), with $C_{miss} = 10$, $C_{FA} = 1$, and $P_{target} = 0.01$.

The out-domain training dataset includes 1,115 male speakers with 13,380 sessions, 1,231 female speakers with 14,772 sessions telephone data selected from Switchboard I, II phase I, II, III corpora. The in-domain dataset includes the telephone data collected from NIST 2004, 2005 and 2006 SRE corpora in total 1034 male speakers with 12,448 sessions and 1,286 female speakers with 14,621 sessions. It is apparent that the adaptation data close to the target region improve the overall system performance. Hence, in this paper, we used simple testing statistics [18] to select the speakers close to the mean of the adaptation dataset. After the selection process, a total number of 711 male speakers with 7,001 sessions and 971 female speakers with 9,281 sessions were used for adaptation.

We used 13-dimensional feature-warped MFCCs with Δ and $\Delta\Delta$ coefficients. Two gender dependent 512 component out-domain UBMs were used to calculate the Baum-Welch statistics for i-vector extraction. The total-variability subspaces of dimension $R_w = 500$ were also trained solely on out-domain data. After compensating the domain mismatch using DMM approach, the DICN transformation was applied to reduce the domain mismatch further in the i-vector subspace. For LDA subspace training, 150 eigenvectors were selected from 500, based on highest eigenvalues. Later, i-vectors were projected into LDA subspace for session and dimensionality reduction. The GPLDA subspaces were trained using length normalized i-vectors and 120 best eigenvoices (N_1) were selected for speaker subspace training. For score normalization, we applied S-normalization technique and created two datasets from in-domain and out-domain datasets to calibrate the domain dependent score normalization performances.

Table 1: *Comparison of PLDA speaker verification on the common set of the NIST-2008 short2-short3 and NIST-2010 core-core evaluation conditions. GPLDA and score normalization are trained using both in-domain and out-domain data.*

GPLDA training	**Score normalization**	**NIST-2008** EER	**NIST-2008** DCF	**NIST-2010** EER	**NIST-2010** DCF$_{old}$
In-domain	In-domain	**3.38%**	**0.0162**	**4.80%**	**0.0207**
	Out-domain	3.62%	0.0177	5.08%	0.0217
Out-domain	In-domain	3.85%	0.0183	4.96%	0.0216
	Out-domain	4.70%	0.0230	5.68%	0.0268

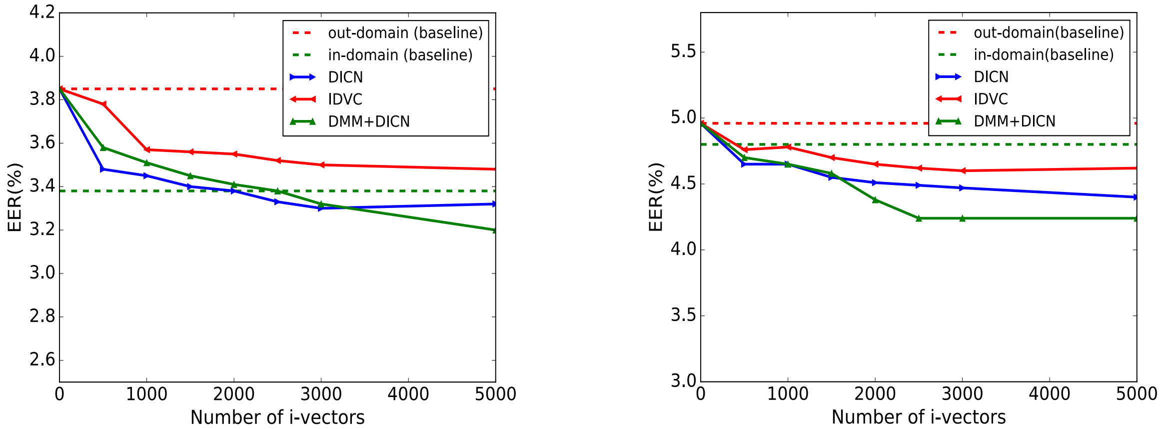
Table 2: *Comparison of different domain adaptation techniques for out-domain GPLDA speaker verification using both labeled and unlabeled data for adaptation.*

Adapt data	**Approach**	**NIST-2008** EER	**NIST-2008** DCF	**NIST-2010** EER	**NIST-2010** DCF$_{old}$
Unlabeled	Out-domain GPLDA	3.85%	0.0183	4.96%	0.0216
	IDVC [19]	3.47%	0.0171	4.62%	0.0200
	DICN [14]	3.29%	0.0154	4.32%	0.0178
	DMM	3.56%	0.0178	4.48%	0.0201
	DMM+DICN	**3.20%**	**0.0149**	**4.24%**	**0.0176**
Labeled	Pooled GPLDA	2.80%	0.0134	3.67%	0.0161
	IDVC (Pooled) [19]	3.12%	0.0132	3.81%	0.0157
	DICN (Pooled) [14]	2.96%	0.0132	3.64%	0.0155
	DMM(Pooled)	2.72%	0.0126	3.61%	0.0153
	DMM+DICN(Pooled)	**2.68%**	**0.0123**	**3.39%**	**0.0143**

6. Results and Discussion

Table 1 shows the performance gap between the in-domain and out-domain GPLDA systems due to the mismatch between training and evaluation data. Experimental results also show that the performance degrades when the score normalization is trained on different domain data. The best result is achieved when GPLDA and score normalization are trained on in-domain data. Since the score normalization trained on in-domain data performs better than out-domain data, the rest of the experiments in this paper use in-domain data for score normalization.

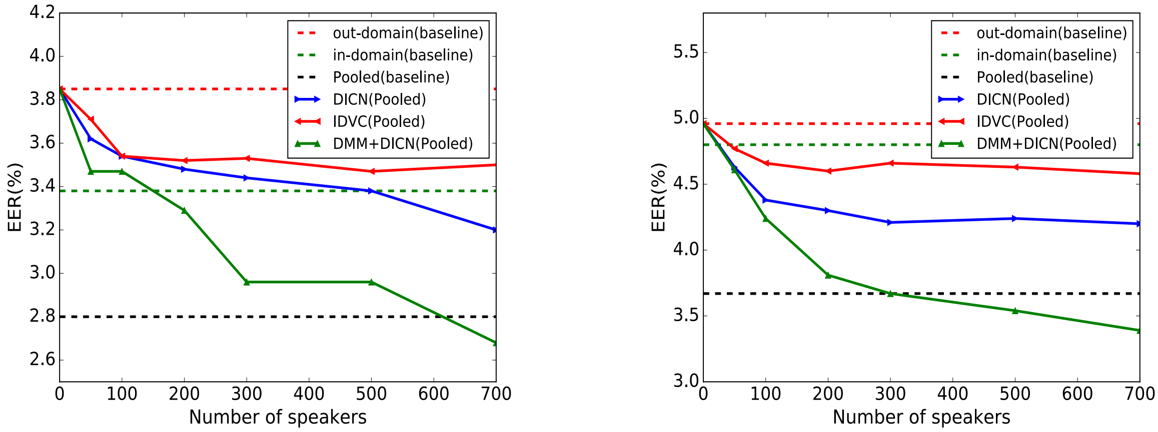
Table 2 compares the performance of different domain adaptation techniques for out-domain GPLDA speaker verification with unlabeled and labeled in-domain adaptation data. The labeled in-domain data are used for domain adaptation as well as pooled with out-domain data for GPLDA training. Experimental results show that the DMM approach is successful in compensating the domain mismatch and achieves 7.5% improvement compared to out-domain baseline system. Although, the DICN approach performs better than DMM approach individually, but combining them together (DMM+DICN) outperforms all other domain adaptation techniques significantly. This suggests that the domain mismatch that can not be modeled with DMM training could be further reduced in the DICN subspace. Subsequently, with the availability of labeled data, the DMM (Pooled) approach outperforms other techniques and achieves at least 27.2% improvement in EER over out-domain baseline and 1.6% improvement in EER over pooled baseline systems. This signifies that in presence of labeled data, the DMM approach



(a) NIST-2008 *short2-short3*

(b) NIST-2010 *core-core*

Figure 1: *The performance comparison of different domain adaptation techniques against the in-domain and out-domain GPLDA baselines by limiting the number of unlabeled i-vectors for (a) NIST 2008 SRE short2-short3 condition and (b) NIST 2010 SRE core-core condition.*



(a) NIST-2008 *short2-short3*

(b) NIST-2010 *core-core*

Figure 2: *The performance comparison of different domain adaptation techniques against the in-domain, out-domain and pooled GPLDA baselines by limiting the number speakers for (a) NIST 2008 SRE short2-short3 condition and (b) NIST 2010 SRE core-core condition.*

captures more domain specific information to compensate the domain mismatch more accurately from the training data. Further improvement is achieved with the combined DMM+DICN approach, gaining at least 14.5% improvement in EER with unlabeled data and 30.4% improvement in EER with labeled data over out-domain baseline system.

We also reduced the adaptation data for domain mismatch compensation training to investigate the data scarce conditions. Figure 1 and 2 show the performance of different domain adaptation techniques with limited unlabeled and labeled adaptation data, respectively. Initially, with a very small amount of data, the DICN approach performs relatively better than IDVC approach, but performance reaches the plateau after using a certain amount of adaptation data. In overall, the combined DMM+DICN gives the best performance compared to other adaptation techniques. For NIST 2008 *telephone-telephone* condition, DICN performs better than combined approach when a small number of i-vectors are available. However, Figure 1 shows that DMM+DICN combined approach gives the best performance when more than 4,000 unlabeled in-domain i-vectors are available for adaptation. Also, with this combined approach, only 2,000 unlabeled i-vectors are required to outperform the in-domain baseline GPDLA system for both NIST 2008 and NIST 2010 evaluations. From Figure 2, it is evident that DMM+DICN (Pooled) approach performs considerably better than other domain adaptation techniques and requires only small number of in-domain speakers (200 speakers) for adaptation to outperform the in-domain baseline system. Also, this approach requires only 600 in domain speakers to perform better than the pooled GPLDA system.

7. Conclusion

In this paper, we presented DMM approach to model and compensate the domain mismatch from the out-domain i-vectors to improve the speaker verification performance. Experiments were carried out on both NIST-2008 and NIST-2010 SRE corpora to demonstrate the domain adaptation performance. Results showed that DMM approach was successful in compensating the domain mismatch, and further performance improvement was achieved by combining DMM with DICN to compensate the domain mismatch further in the i-vector subspace. This combined DMM+DICN approach successfully achieved at least 30.4% performance gain when only a small amount of labeled data is available for domain adaptation. Experimental studies using limited in-domain data showed that the combined approach required only small amount of i-vectors (2,000 unlabeled or 200 labeled speakers) to perform better than the out-domain GPLDA system.

8. References

[1] Patrick Kenny, "Bayesian speaker verification with heavy tailed priors," in *Speaker and Language Recognition Workshop (IEEE Odyssey)*, 2010.

[2] "Speaker and language recognition summer workshop, Tech. rep., NIST (2013)," URL online: http://www.clsp.jhu.edu/workshops/archive/ws13-summer-workshop/groups/spk-13/.

[3] Daniel Garcia-Romero and Alan McCree, "Supervised domain adaptation for i-vector based speaker recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, 2014, pp. 4047–4051.

[4] Qiongqiong Wang, Hitoshi Yamamoto, and Takafumi Koshinaka, "Domain adaptation using maximum likelihood linear transformation for PLDA-based speaker verification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 5110–5114.

[5] Qingyang Hong, Jun Zhang, Lin Li, Lihong Wan, and Feng Tong, "A transfer learning method for PLDA-based speaker verification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 5455–5459.

[6] Jesús Villalba and Eduardo Lleida, "Bayesian adaptation of PLDA based speaker recognition to domains with scarce development data," in *The Speaker and Language Recognition Workshop (IEEE Odyssey)*, 2012.

[7] Hagai Aronowitz, "Inter dataset variability compensation for speaker recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, 2014, pp. 4002–4006.

[8] Elliot Singer and Douglas A Reynolds, "Domain mismatch compensation for speaker recognition using a library of whiteners," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2000–2003, 2015.

[9] Daniel Garcia-Romero, Alan McCree, Stephen Shum, Niko Brümmer, and Carlos Vaquero, "Unsupervised domain adaptation for i-vector speaker recognition," in *The Speaker and Language Recognition Workshop (IEEE Odyssey)*, 2014, pp. 260–264.

[10] Daniel Garcia-Romero, Xiaohui Zhang, Alan McCree, and Daniel Povey, "Improving speaker recognition performance in the domain adaptation challenge using deep neural networks," in *IEEE Spoken Language Technology Workshop (SLT)*, 2014, pp. 378–383.

[11] Stephen Shum, Najim Dehak, Reda Dehak, and James R Glass, "Unsupervised speaker adaptation based on the cosine similarity for text-independent speaker verification.," in *The Speaker and Language Recognition Workshop (IEEE Odyssey)*, 2010, p. 16.

[12] Jesús Villalba and Eduardo Lleida, "Unsupervised adaptation of PLDA by using variational bayes methods," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, 2014, pp. 744–748.

[13] Ondrej Glembek, Jeff Ma, Pavel Matejka, Bing Zhang, Oldrich Plchot, Lukas Burget, and Spyros Matsoukas, "Domain adaptation via within-class covariance correction in i-vector based speaker recognition systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, 2014, pp. 4032–4036.

[14] Md Hafizur Rahman, Ahilan Kanagasundaram, David Dean, and Sridha Sridharan, "Dataset-invariant covariance normalization for out-domain PLDA speaker verification," in *Interspeech*, September 2015, pp. 1017–1021.

[15] Waad Ben Kheder, Driss Matrouf, Pierre-Michel Bousquet, Jean-François Bonastre, and Moez Ajili, "Robust speaker recognition using MAP estimation of additive noise in i-vectors space," in *International Conference on Statistical Language and Speech Processing*. Springer, 2014, pp. 97–107.

[16] Najim Dehak, Patrick Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[17] Daniel Garcia-Romero and Carol Y Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems.," in *Interspeech*, 2011, pp. 249–252.

[18] Zhidong Bai and Hewa Saranadasa, "Effect of high dimension: by an example of a two sample problem," *Statistica Sinica*, pp. 311–329, 1996.

[19] Ahilan Kanagasundaram, David Dean, and Sridha Sridharan, "Improving out-domain PLDA speaker verification using unsupervised inter-dataset variability compensation approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, 2015.