Dataset-Invariant Covariance Normalization for Out-domain PLDA Speaker Verification

Md Hafizur Rahman[1], *Ahilan Kanagasundaram*[2], *David Dean*[3], *Sridha Sridharan*[4]

Speech and Audio Research Laboratory
Queensland University of Technology, Brisbane, Australia
{m20.rahman[1], a.kanagasundaram[2], s.sridharan[4]}@qut.edu.au, ddean@ieee.org[3]

Abstract

In this paper we introduce a novel domain-invariant covariance normalization (DICN) technique to relocate both in-domain and out-domain i-vectors into a third dataset-invariant space, providing an improvement for out-domain PLDA speaker verification with a very small number of unlabelled in-domain adaptation i-vectors. By capturing the dataset variance from a global mean using both development out-domain i-vectors and limited unlabelled in-domain i-vectors, we could obtain domain-invariant representations of PLDA training data. The DICN-compensated out-domain PLDA system is shown to perform as well as in-domain PLDA training with as few as 500 unlabelled in-domain i-vectors for NIST-2010 SRE and 2000 unlabelled in-domain i-vectors for NIST-2008 SRE, and considerable relative improvement over both out-domain and in-domain PLDA development if more are available.

Index Terms: speaker verification, PLDA, DICN, domain adaptation

1. Introduction

In the past few years extensive research has been conducted in the field of speaker verification. Numerous methods have been proposed, like joint factor analysis (JFA) [1] and i-vector [2] based subspace modelling techniques, that have resulted in excellent speaker verification performance. But most techniques have only been investigated in relatively clean environments, with huge amounts of 'in-domain' development data. However, if we use this clean data for development of real world applications, it would produce poor performance, because of many factors that are not always considered in clean development data. One of the key reason is the mismatch between development and evaluation dataset.

The performance variation due to cross-domain speaker verification development and evaluation was first addressed at the Summer Workshop at Johns Hopkins University (JHU) held in 2013 [3]. Results presented in that workshop clearly showed the performance gap between in-domain and out-domain development for speaker verification. This task was deemed the 'Domain Adaptation Challenge' (DAC) at that workshop.

In response to this poor cross-domain speaker verification performance, Garcia-Romero *et al.* [4] found that training UBMs and total-variability matrices on in-domain or out-domain data have very limited effect on overall performance, but the effect on PLDA parameters were more pronounced. They investigated four adaptation techniques for supervised domain adaptation of PLDA parameters: fully Bayesian adaptation, approximate MAP, weighted likelihood and SPLDA parameter interpolation. Each of these techniques performed very similarly. Villalba *et al.* [5] introduced a variational Bayesian technique for adapting PLDA models from labeled out-domain to unlabeled in-domain data. Recently, Garcia-Romero *et al.* also introduced an agglomerative hierarchical clustering (AHC) method to cluster unlabeleld in-domain data for domain adaptation. To compensate the dataset shift in i-vector space Aronowitz [6] introduced an inter-dataset variability compensation (IDVC) technique based on nuisance attribute projection (NAP). Glembek *et al.* [7] proposed within-speaker covariance correction (WCC) and extended unsupervised adaptation of the LDA matrix to compensate the mismatch between training and testing datasets. Recently, Kanagasundaram *et al.* [8] introduced an improved IDVC technique, where dataset variability is captured using difference between out-domain i-vectors and average of in-domain i-vectors.

In this paper, a novel dataset invariant covariance normalization (DICN) approach is introduced to compensate the mismatch between in-domain and out-domain dataset in the i-vector space. Instead of capturing the mismatch directly between out-domain and in-domain data [8], we captured the mismatch as compared to the global mean i-vector. In this approach we used a set of unlabelled in-domain i-vectors and captured the mismatch using the difference between all i-vectors (in-domain and out-domain) and the global mean i-vector.

The rest of the paper is structured as follows: Section 2 details the i-vector feature extraction techniques. Section 3 details the DICN approach. Section 4 explains the linear discriminant analysis (LDA), and Section 5 presents GPLDA based speaker verification system. The experimental setup and corresponding results are given in Section 6 and Section 7. Finally, Section 8 concludes the paper.

2. I-vector based speaker verification

Single subspace-based i-vector speaker verification was first proposed by Dehak *et al.* [2]. This approach was inspired by his previous work, finding speaker discriminant information was lost in the discarded channel space of the earlier joint factor analysis (JFA) technique [9]. Unlike the JFA approach, in i-vector feature extraction the GMM super-vectors are represented in a single subspace called total-variability subspace. Both speaker and channel dependent GMM super-vector in i-vector can be represented by,

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

where $\boldsymbol{m}$ is the speaker and session independent UBM super-vector, $\boldsymbol{T}$ is a low rank matrix. $\boldsymbol{w}$ is total variability factor