Available online at www.sciencedirect.com

ScienceDirect

Computer Speech & Language 47 (2018) 240–258

www.elsevier.com/locate/csl

Improving PLDA speaker verification performance using domain mismatch compensation techniques☆

Md Hafizur Rahman*, Ahilan Kanagasundaram, Ivan Himawan, David Dean, Sridha Sridharan

Speech and Audio Research Lab, SAIVT, Queensland University of Technology, Australia

Received 9 October 2016; received in revised form 25 July 2017; accepted 7 August 2017
Available online 18 August 2017

Abstract

The performance of state-of-the-art i-vector speaker verification systems relies on a large amount of training data for probabilistic linear discriminant analysis (PLDA) modeling. During the evaluation, it is also crucial that the target condition data is matched well with the development data used for PLDA training. However, in many practical scenarios, these systems have to be developed, and trained, using data that is often outside the domain of the intended application, since the collection of a significant amount of in-domain data is often difficult. Experimental studies have found that PLDA speaker verification performance degrades significantly due to this development/evaluation mismatch. This paper introduces a domain-invariant linear discriminant analysis (DI-LDA) technique for out-domain PLDA speaker verification that compensates domain mismatch in the LDA subspace. We also propose a domain-invariant probabilistic linear discriminant analysis (DI-PLDA) technique for domain mismatch modeling in the PLDA subspace, using only a small amount of in-domain data. In addition, we propose the sequential and score-level combination of DI-LDA, and DI-PLDA to further improve out-domain speaker verification performance. Experimental results show the proposed domain mismatch compensation techniques yield at least 27% and 14.5% improvement in equal error rate (EER) over a pooled PLDA system for *telephone-telephone* and *interview-interview* conditions, respectively. Finally, we show that the improvement over the baseline pooled system can be attained even when significantly reducing the number of in-domain speakers, down to 30 in most of the evaluation conditions.

Keywords: Speaker verification; I-vector; Domain mismatch compensation; DI-LDA; DI-PLDA; DI-PLDA[DI-LDA]; Score fusion

1. Introduction

Over the past few years, speaker verification technology has evolved rapidly, especially after the introduction of joint factor analysis (JFA) by Kenny (2005). JFA allowed speaker and channel variability modeling explicitly from high-dimensional Gaussian mixture model (GMM) super-vectors by considering GMM super-vectors as a linear

☆ This paper has been recommended for acceptance by Roger K. Moore.

* Corresponding author.
E-mail address: m20.rahman@qut.edu.au (M.H. Rahman), a.kanagasundaram@qut.edu.au (A. Kanagasundaram), i.himawan@qut.edu.au (I. Himawan), ddean@ieee.org (D. Dean), s.sridharan@qut.edu.au (S. Sridharan).

http://dx.doi.org/10.1016/j.csl.2017.08.001

combination of speaker and channel components. Subsequently, JFA advanced into an i-vector extraction technique (Dehak et al., 2011), where a low dimensional total-variability space is trained to represent both speaker and channel models together, instead of modeling speaker and channel variability separately. The fundamental idea of this single space representation was to capture some explicit speaker dependent information previously lost in the channel space under JFA. To reduce the channel effect on the i-vectors, different channel compensation techniques were introduced, such as within-class covariance normalization (WCCN), linear discriminant analysis (LDA) and nuisance attribute projection (NAP) (Dehak et al., 2011). Kenny (2010) proposed a probabilistic linear discriminant analysis (PLDA) technique for modeling speaker and session variability in the i-vector subspace to further improving the performance of speaker verification system. Originally, PLDA was used with a heavy-tailed Student's t-distribution assumption, but later, Garcia-Romero and Espy-Wilson (2011) introduced length-normalized Gaussian PLDA (GPLDA), which was computationally more efficient compared to heavy-tailed PLDA (HTPLDA). Recently, a combination of a deep neural network (DNN) with i-vectors has become very popular in speaker recognition task, where posteriors of the DNN senones are extracted and appended with mel-frequency cepstral coefficients (MFCC) features (McLaren et al., 2015; Sadjadi et al., 2016; Richardson et al., 2015).

The success of the state-of-the-art i-vector based PLDA speaker verification systems are dependent on the volume of speech data as well as the domain information to train them successfully for low error rates. Unfortunately, it is not always feasible to collect large amount of target-domain speech data to develop practical speaker verification systems. Automatic speaker verification system performance degrades substantially when trained on out-domain data, due to the domain mismatch between out-domain training data and the data used in the target applications. Similar issues have been investigated in the Speaker and Language Recognition Workshop at Johns Hopkins University (JHU, 2013). The in-domain dataset collected from National Institute of Standards and Technology (NIST) telephone data and out-domain dataset collected from LDC Switchboard corpus (SWB) telephone data were used in this challenge. The preliminary findings presented at the workshop showed that domain mismatch contributes roughly 15−40% performance degradation in the PLDA speaker verification systems. In this paper, we address the problem of improving out-domain PLDA speaker verification performance using a number of domain mismatch compensation techniques.

2. Related work

Different techniques have been proposed recently to improve the state-of-the-art PLDA speaker verification performance, when trained on out-domain data. These techniques can be broadly categorized into supervised and unsupervised adaptation methods. For supervised adaptation, a small amount of in-domain data with speaker labels are available. Villalba and Lleida (2012) investigated a Bayesian adaptation technique for adapting PLDA parameters to limited supervised target domain data. They used a fully Bayesian approach and a variational approximation to compute the intractable posterior using conjugate priors. Due to the adaptation of the channel matrix, they found good performance in the domain with limited supervised development data. Garcia-Romero and McCree (2014) concluded that training UBMs and total variability subspaces on out-domain data do not change the system performance significantly. Garcia-Romero *et al.* proposed four supervised PLDA parameters adaptation techniques: fully-Bayesian adaptation, approximate maximum a posteriori (MAP) adaptation, weighted likelihood and parameter interpolation (Garcia-Romero and McCree, 2014). All of them performed very similarly in compensating domain mismatch from PLDA subspace. Wang et al. (2016) proposed a maximum likelihood linear transformation (MLLT) approach to transfer features from the target domain to the source domain. They introduced two PLDA-MLLT techniques to estimate the transfer parameters and used expectation maximization (EM) to estimate the new adapted PLDA parameters. These techniques outperformed the linear PLDA parameter interpolation technique. Hong et al. (2016) proposed a transfer learning method based on Bayesian joint probability, where they used Kullback-Leibler (KL) divergence approach to maximize a new optimization function for PLDA training to share information between the domains. This approach also produced relatively-better performance than the traditional PLDA and interpolated PLDA approaches. Aronowitz (2014b) proposed an inter dataset variability compensation (IDVC) approach to minimize the domain mismatch in the i-vector subspace using nuisance attribute projection (NAP). He partitioned the large out-domain dataset into 12 small subsets, and formulated an IDV subspace using the principal component analysis (PCA) technique, spanned by the 12 centers of those subsets. This technique successfully compensated dataset shift in the i-vector subspace in the context of the domain adaptation challenge. Singer and Reynolds (2015)