Odyssey 2012
The Speaker and Language Recognition Workshop
25-28 June 2012, Singapore

PLDA based Speaker Verification with Weighted LDA Techniques

Ahilan Kanagasundaram, David Dean, Sridha Sridharan, Robbie Vogt

Speech and Audio Research Laboratory
Queensland University of Technology, Brisbane, Australia
{a.kanagasundaram, d.dean, s.sridharan, r.vogt}@qut.edu.au

Abstract

This paper investigates the use of the dimensionality-reduction techniques weighted linear discriminant analysis (WLDA), and weighted median fisher discriminant analysis (WMFD), before probabilistic linear discriminant analysis (PLDA) modeling for the purpose of improving speaker verification performance in the presence of high inter-session variability. Recently it was shown that WLDA techniques can provide improvement over traditional linear discriminant analysis (LDA) for channel compensation in i-vector based speaker verification systems. We show in this paper that the speaker discriminative information that is available in the distance between pair of speakers clustered in the development i-vector space can also be exploited in heavy-tailed PLDA modeling by using the weighted discriminant approaches prior to PLDA modeling. Based upon the results presented within this paper using the NIST 2008 Speaker Recognition Evaluation dataset, we believe that WLDA and WMFD projections before PLDA modeling can provide an improved approach when compared to uncompensated PLDA modeling for i-vector based speaker verification systems.

1. Introduction

I-vector-based speaker verification has recently become the state of the art of speaker verification, providing superior performance when compared to joint factor analysis (JFA) approach [1]. Rather than taking the JFA approach of modeling speaker and channel variability spaces separately, the i-vector approach forms a low-dimensional, total-variability space that models both speaker and channel variability together. Unlike JFA, where factor analysis is used to generate a discriminative model, the i-vector approach uses similar factor analysis techniques as a feature extractor, creating an intermediate speaker representation between the high dimensional Gaussian mixture model (GMM) super-vector and traditional low dimensional acoustic feature representations [1]. As the channel variation is included within the total variability space, i-vector features are often combined with channel compensation techniques to attenuate channel variation in the i-vector space. The choice of channel compensation techniques have become a very active area of research, with initial research focusing on the use of linear discriminant analysis (LDA) followed by within-class covariance normalization (WCCN), as proposed by Dehak *et al.* [2]. Recently, this approach was extended by McLaren and van Leeuwen [3] who proposed a new LDA-based approach, source-normalized LDA (SN-LDA), which improves the i-vector speaker representation in both mismatched conditions and conditions for which limited hyperparameter developmental speech resources are available. This work has been futher extended by Kanagasundaram *et al.*, by investigating new channel compensation approaches of weighted LDA (WLDA) and source-normalized weighted LDA (SN-WLDA) [4], and these were found to achieve further improvement over both the non-weighted LDA and SN-LDA techniques.

Recently these low dimensional i-vector features were extended with a probabilistic linear discriminant analysis (PLDA) approach to model speaker and channel part within the i-vector space, and this has been shown to provide improved speaker verification performance to the initial i-vector approach [5, 6, 7]. This PLDA technique was originally proposed by Price *et al.* [8] for face recognition, and was adapted to i-vectors for speaker verification by Kenny *et al.* [5, 6, 7]. In his original paper, Kenny investigated two generative approaches to forming the PLDA models: Gaussian PLDA (GPLDA) and heavy-tailed PLDA (HTPLDA) [5]. Kenny found that HTPLDA achieved significant improvement over GPLDA, concluding that i-vector features are better modeled by heavy-tailed distribution due to the frequent presence of outliers in the i-vector space. More recently Matejka *et al.* have investigated dimensionality reduction using LDA before PLDA modeling [9], and achieved an improvement on *telephone-telephone* (*enrolment-verification*) conditions. However this approach of transforming the i-vector space before PLDA modeling has not yet been investigated under mismatched conditions where *enrolment* and *verification* conditions are not matched. More importantly, the investigation of more advanced channel compensation techniques such as WLDA, median fisher discriminator (MFD), and weighted MFD (WMFD) would be of considerable value to improving PLDA-based speaker verification systems.

The advantages of LDA-based approaches is that a higher dimensional i-vector feature can be projected into a much lower dimensional space with minimal loss of discriminantive ability, as the ratio of between-speaker and within-speaker variations is maximized. The between-speaker variation normally depends on speaker's characteristics, but the within-speaker variation is much more dependent on the choice of microphone, the acoustic environment, transmission channels and day-to-day differences within a speakers voice. The full potential of using LDA-based approaches with i-vector speaker verification system is not realized with traditional LDA due to the large channel variation and the heavy-tailed behavior of i-vector distributions. We investigate in this paper if channel compensation using LDA, WLDA, MFD, and WMFD can provide superior performance for HTPLDA based speaker verification over non-channel compensated approaches.

This paper is structured as follows: Section 2 gives a brief introduction to the process of PLDA based speaker verification and also introduces i-vector feature extraction, dimensionality reduction techniques, PLDA modeling and scoring. Section 3 describes the methodology of the experiments conducted in this paper, and results and corresponding discussions are given in Section 4. Section 5 concludes the paper.