Improving PLDA Speaker Verification using WMFD and Linear-weighted Approaches in Limited Microphone Data Conditions

Ahilan Kanagasundaram, David Dean and Sridha Sridharan

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

Abstract

This paper proposes the addition of a weighted median Fisher discriminator (WMFD) projection prior to length-normalised Gaussian probabilistic linear discriminant analysis (GPLDA) modelling in order to compensate the additional session variation. In limited microphone data conditions, a linear-weighted approach is introduced to increase the influence of microphone speech dataset. The linear-weighted WMFD-projected GPLDA system shows improvements in EER and DCF values over the pooled LDA- and WMFD-projected GPLDA systems in *interview-interview* condition as WMFD projection extracts more speaker discriminant information with limited number of sessions/ speaker data, and linear-weighted GPLDA approach estimates reliable model parameters with limited microphone data.
Index Terms: speaker verification, i-vectors, GPLDA, WMFD, linear-weighted

1. Introduction

A significant amount of speech data is required to develop a robust speaker verification system, especially in the presence of large intersession variability [1]. However, it is often difficult to acquire a sufficient number of sessions for each individual speaker for developing robust background models in many real-world environments, limiting the availability of speaker verification technology for many everyday applications. A significant example of this problem is the relative scarcity of microphone speech data available across the many NIST Speaker Recognition Evaluation (SRE) databases [2, 3], which have, at least until more recent evaluations, focussed largely on collecting large quantities of telephone speech.

Speaker verification is a data-driven research field, and it has clearly been established that the development of state-of-the-art speaker verification systems require a significant volume of speech data covering multiple sessions across a large number of speakers [1]. However, the volume of data required to adequately model the background behaviour of speaker models is not always available, particularly in new environments. Recently, we have analysed the linear discriminant analysis (LDA) projected Gaussian probabilistic linear discriminant analysis (GPLDA) speaker verification system with limited development data, and found that when the number of sessions/speaker are reduced, the speaker verification performance is considerably affected [4]. As an alternative approach to LDA projection, we have also previously introduced the median Fisher discriminator (MFD) and a weighted variant (WMFD) to show better speaker discriminative performance from limited-session development data than the mean-centroid approach of LDA [4].

In addressing the disparate microphone and telephone data sources available in the NIST evaluations, researchers have shown that pooling the telephone and microphone speech data is the best approach for the development of GPLDA [5, 6] speaker verification systems. In our recent work, we have introduced a linear-weighted approach to effectively model the GPLDA parameters proportionally from telephone and microphone speech data [7] that has shown promise in limited development session conditions.

In this paper, initially a LDA-projected GPLDA speaker verification system was analysed with limited development data to investigate the effect on speaker verification performance. This approach is then compared to the alternative, WMFD-projected linear-weighted GPLDA approach, to show an improvement in speaker verification performance for limited microphone development sessions. In our GPLDA speaker verification system, telephone speakers (of which we are developing across 1286 female and 1034 male speakers) with more than 10 sessions and a limited number of microphone speakers (100 female and 83 male speakers) with more than 15 session were used for GPLDA modelling. To demonstrate the effect of limited sessions during development, we have restricted both the telephone and microphone speech speakers to 7 sessions per speaker.

This paper is structured as follows: Section 2 outlines a typical state-of-the-art GPLDA speaker verification system, and Section 3 gives a brief overview of dimensionality reduction approaches, including LDA and WMFD. Section 4 details the GPLDA model-parameter estimation techniques for scarce microphone speech. The experimental protocol and corresponding results are given in Section 5 and Section 6, and Section 7 concludes the paper.

2. GPLDA Speaker Verification

2.1. I-vectors

I-vectors represent a Gaussian mixture model (GMM) mean super-vector by a single total-variability subspace. This single-subspace approach was motivated by the discovery that the channel space of the earlier, related JFA technique contained valuable speaker-discriminant information [8]. An i-vector speaker-and-channel-dependent GMM super-vector $\boldsymbol{\mu}$ can be represented by,

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

where $\mathbf{m}$ is a universal background model (UBM) mean super-vector trained over a large development set and $\mathbf{T}$ is a low-rank total-variability matrix. The total-variability factors ($\mathbf{w}$) are the i-vectors, and are normally distributed with parameters *N(0,1)*. Extracting an i-vector from the total-variability