Short Utterance PLDA Speaker Verification using SN-WLDA and Variance Modelling Techniques

Ahilan Kanagasundaram, David Dean, 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 a combination of source-normalized weighted linear discriminant analysis (SN-WLDA) and short utterance variance (SUV) PLDA modelling to improve the short utterance PLDA speaker verification. As short-length utterance i-vectors vary with the speaker, session variations and phonetic content of the utterance (utterance variation), a combined approach of SN-WLDA projection and SUV PLDA modelling is used to compensate the session and utterance variations. Experimental studies have found that a combination of SN-WLDA and SUV PLDA modelling approach shows an improvement over baseline system (WCCN[LDA]-projected Gaussian PLDA (GPLDA)) as this approach effectively compensates the session and utterance variations.

Index Terms: speaker verification, session variation, utterance variation, LDA, SN-WLDA

1. Introduction

A significant amount of speech is required for speaker model enrolment and verification, especially in the presence of large intersession variability, which has limited the widespread use of speaker verification technology in everyday applications. Reducing the amount of speech required for development, enrolment and verification while obtaining satisfactory performance has been the focus of a number of recent studies in state-of-the-art speaker verification design, including joint factor analysis (JFA), i-vectors, probabilistic linear discriminant analysis (PLDA) and support vector machines (SVM) [1, 2, 3, 4, 5, 6]. Recently, Kenny *et al.* [7], have investigated how to quantify the uncertainty associated with the i-vector extraction process and propagate it into a PLDA classifier. Continuous research on this field has been ongoing to address the robustness of speaker verification technologies under such conditions.

The total-variability, or *i-vector*, approach has risen to prominence as the de-facto standard in recent state-of-the-art speaker verification systems, due to its intrinsic capability to map an utterance to a single low-dimensional i-vector, turning a complex high-dimensional speaker recognition problem into a low-dimensional classical pattern recognition one. However, i-vectors extracted from different durations should not be considered equal in reliability concerns. Moreover, long utterance i-vectors vary with speaker and session variations whereas short utterance i-vectors contain speaker, session and utterance variations, and these session and utterance variations need to be compensated in short utterance speaker verification.

As the session variability is included within the i-vector space, PLDA approach is commonly used to model speaker and session variations [8, 9]. In recent times, prior to the PLDA modelling, linear discriminant analysis (LDA) followed by within-class covariance normalization (WCCN) (WCCN[LDA]) session compensation approach is applied to compensate the additional session variation and reduce the computational complexness [10]. Recently, we have introduced the short utterance variance normalisation (SUVN) and short utterance variance (SUV) modelling to cosine similarity scoring (CSS) i-vector and PLDA speaker verification systems to compensate the session and utterance variations [11].

The main aim of this paper is to find a method that effectively compensates the session and utterance variations. Previously, we have found that source-normalised weighted LDA (SN-WLDA) followed by WCCN (WCCN[SN-WLDA])-projected Gaussian PLDA (GPLDA) system effectively compensates the session variation than standard WCCN[LDA]-projected GPLDA system in long utterance evaluation conditions [12]. Recently, it was also found that WCCN projection doesn't provide any advantage to PLDA speaker verification as PLDA models the intra-speaker variance itself [11]. In this paper, initially SN-WLDA-projected GPLDA system is studied with short utterance evaluation conditions. Subsequently, a combination of SN-WLDA and SUV modelling approach is introduced to PLDA speaker verification to effectively compensate the session and utterance variations.

This paper is structured as follows: Section 2 details the i-vector feature extraction techniques, and Section 3 explains how short utterance variance is added to i-vector features. Section 4 gives a brief details of SN-WLDA and SUV modelling approaches. Section 5 explains the GPLDA based speaker verification system. The experimental protocol and corresponding results are given in Section 6 and Section 7. Section 8 concludes the paper.

2. I-vector feature extraction

I-vectors represent the GMM super-vector by a single total-variability subspace. This single-subspace approach was motivated by the discovery that the channel space of JFA contains information that can be used to distinguish between speakers [13]. An i-vector speaker and channel dependent GMM super-vector can be represented by,

$$\boldsymbol{\mu} \quad = \quad \mathbf{m} + \mathbf{T}\mathbf{w}, \qquad (1)$$

where $\mathbf{m}$ is the same universal background model (UBM) super-vector used in the JFA approach 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 subspace is es-