Improving the PLDA based Speaker Verification in Limited Microphone Data Conditions

A. Kanagasundaram[⋆]*, D. Dean*[⋆]*, J. Gonzalez-Dominguez*[†]*,*
S. Sridharan[⋆]*, D. Ramos*[†]*, J. Gonzalez-Rodriguez*[†]

[⋆] Speech Research Laboratory, Queensland University of Technology, Australia
[†] ATVS Biometric Recognition Group, Universidad Autonoma de Madrid, Spain
[⋆]{a.kanagasundaram, d.dean, s.sridharan}@qut.edu.au,
[†] {javier.gonzalez, daniel.ramos, joaquin.gonzalez} @uam.es

Abstract

A significant amount of speech data is required to develop a robust speaker verification system, but it is difficult to find enough development speech to match all expected conditions. In this paper we introduce a new approach to Gaussian probabilistic linear discriminant analysis (GPLDA) to estimate reliable model parameters as a linearly weighted model taking more input from the large volume of available telephone data and smaller proportional input from limited microphone data. In comparison to a traditional pooled training approach, where the GPLDA model is trained over both telephone and microphone speech, this linear-weighted GPLDA approach is shown to provide better EER and DCF performance in microphone and mixed conditions in both the NIST 2008 and NIST 2010 evaluation corpora. Based upon these results, we believe that linear-weighted GPLDA will provide a better approach than pooled GPLDA, allowing for the further improvement of GPLDA speaker verification in conditions with limited development data.

Index Terms: speaker verification, i-vector, total-variability, length-normalization, Gaussian PLDA

1. Introduction

A significant amount speech data is required to develop a robust speaker verification system, especially in the presence of large intersession variability. In practice it is feasible to collect substantial amount of telephone data but microphone data is harder to acquire. For example, large amount of telephone speech data is available in the NIST Speaker Recognition Evaluation (SRE) databases; however, microphone speech data is scarce in this data set. In addressing this disparity data sources, researchers have pooled the telephone and microphone data for the development of modern state of the art speaker verification systems such as the Gaussian probabilistic linear discriminant analysis (GPLDA) [3]. In this paper we take a new approach to estimate reliable model parameters as a linear weighted model, taking more input from the large volume of available telephone data and smaller proportional input from limited microphone data.

Joint factor analysis (JFA), as originally proposed by Kenny [4], has evolved as a powerful tool in speaker verification to model the inter-speaker variability and to compensate for channel/session variability in the context of high-dimensional Gaussian mixture model (GMM) super-vectors. Dominguez *et al.* [1] have previously investigated the JFA approach with limited microphone speech data, and have analyzed several approaches, including joining matrices, pooled statistics and scaling statistics to avoid the data scarcity problem.

A few years ago, Dehak *et al.* [5] have proposed a front-end factor analysis technique, termed i-vector, which has evolved from JFA. Rather than taking the JFA approach of modeling a speaker and channel variability space separately, the i-vector approach forms a low-dimensional total-variability space that models both speaker and channel variability. Senoussaoui *et al.* [2] have extended Dehak work, where they have analyzed the i-vector speaker verification approach with microphone speech. They have introduced the concatenated total-variability approach to extract i-vector features from telephone and microphone sources, where total-variability approach is separately trained using telephone and microphone sources and concatenated to form a concatenated total-variability space [2].

Recently, Kenny introduced the PLDA approach to directly model speaker and session variability within the i-vector space [6], and Senoussaoui *et al.* [3] have analyzed the heavy-tailed PLDA (HTPLDA) approach with microphone data conditions. They applied a concatenated total-variability approach to extract useful speaker information from telephone and microphone speech data. However, there have been no investigations into how the length-normalized GPLDA model parameters can be explicitly modeled using both rich telephone and limited microphone speech data.

The main aim of this paper is to explicitly model the GPLDA model parameters using rich telephone and limited microphone sourced speech data in the PDLA model domain. Initially, in the i-vector feature domain, two different types of total-variability approaches, including pooled and concatenated approaches are analyzed to extract the speaker information from telephone microphone speech data. Subsequently, in the PLDA model domain, pooled and linear weighted approaches are investigated to effectively model the GPLDA model parameter from telephone and microphone speech data.

This paper is structured as follows: Section 2 details the i-vector feature extraction techniques. Section 3 gives a brief introduction to the length-normalized GPLDA based speaker verification system, and Section 4 details the GPLDA model parameter estimation methods in scarce microphone speech data. The experimental protocol and corresponding results are given in Section 5 and Section 6. Section 7 concludes the paper.

25 – 29 August 2013, Lyon, France