ELSEVIER

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.
m20.rahman@qut.edu.au, a.kanagasundaram@qut.edu.au, i.himawan@qut.edu.au,
ddean@ieee.org, s.sridharan@qut.edu.au

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 which 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 [1]. 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 combination of speaker and channel components. Subsequently, JFA advanced into an i-vector extraction technique [2], 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) [2]. In [3], Kenny proposed a probabilistic linear discriminant analysis (PLDA) technique for modeling speaker and session vari-