IMPROVING PLDA SPEAKER VERIFICATION WITH LIMITED DEVELOPMENT DATA

Ahilan Kanagasundaram, David Dean and Sridha Sridharan

Speech Research Laboratory, Queensland University of Technology, Australia

{a.kanagasundaram, d.dean, s.sridharan}@qut.edu.au

ABSTRACT

This paper analyses the probabilistic linear discriminant analysis (PLDA) speaker verification approach with limited development data. This paper investigates the use of the median as the central tendency of a speaker's i-vector representation, and the effectiveness of weighted discriminative techniques on the performance of state-of-the-art length-normalised Gaussian PLDA (GPLDA) speaker verification systems. The analysis within shows that the median (using a median fisher discriminator (MFD)) provides a better representation of a speaker when the number of representative i-vectors available during development is reduced, and that further, usage of the pair-wise weighting approach in weighted LDA and weighted MFD provides further improvement in limited development conditions. Best performance is obtained using a weighted MFD approach, which shows over 10% improvement in EER over the baseline GPLDA system on mismatched and *interview-interview* conditions.

Index Terms— Speaker verification, PLDA, WLDA, WMFD

1. INTRODUCTION

Speaker verification has traditionally required a large volume of speech during development and evaluation, particularly in the presence of high intersession variability. However, it can be hard to acquire a sufficient quantity of data in many real-world environments, limiting the suitability of speaker verification for many everyday applications. Recently, a number of interesting techniques have focussed on reducing the amount of speech required during evaluation (covering the enrolment of speaker models, and their verification), but little effort has been put into reducing the volume of speech required to develop new models for deployment into previously unseen environments.

Reducing the amount of speech required during enrolment and verification whilst maintaining satisfactory performance has been the focus in a number of recent studies focused on a variety of speaker verification technologies: joint factor analysis (JFA) [1], support vector machines (SVM) [2], i-vectors [3] and probabilistic linear discriminant analysis (PLDA) [4]. These studies have shown that, across all technologies, the verification performance degrades considerably when very-short utterances ($< 10s$) are used as evaluation data. A number of attempts to compensate for this reduction in performance have been undertaken in the literature. Kenny *et al.* [5] have investigated how to quantify the uncertainty associated with summarising various length utterances down to a constant-length i-vector and demonstrated how to propagated that into the PLDA classifier. An alternative approach was demonstrated by Hasan *et al.* [6] where they found that the duration variability can be modelled as additive noise in the i-vector space, also using a PLDA classifier.

Speaker verification is a data-driven research field, and it has clearly been established that state-of-the-art speaker verification systems require a significant volume of development data covering multiple sessions across a large number of speakers [7]. However, the volume of data required to adequately model the background behaviour of speaker models is not always available, particularly in new environments. In a recent study, the i-vector and PLDA speaker verification systems' performance were analysed when the long- and short-length utterance development data was used for speaker development, where Kanagasundaram *et al.* have found that instead of using the full-length utterance development data, when short-length utterance development data is used for PLDA modelling, speaker verification systems shows a significant improvement [4]. However, there hasn't yet been any detailed investigations on how state-of-the-art PLDA speaker verification copes with limited session development data.

In this paper, initially a LDA-projected Gaussian PLDA (GPLDA) speaker verification system's is analysed with limited development data to investigate the effect on speaker verification performance. An alternative approach to LDA projection, the median Fisher discriminator (MFD) is then introduced to show better speaker discriminative performance from limited channel development data than the mean-centroid approach of LDA. Finally, weighted approaches, where weighting the speakers that are closer to each other to reduce speaker confusion, of LDA (WLDA) and MFD (WMFD) are introduced to provide a further boost in speaker

This paper was based upon research conducted through the Australian Research Council (ARC) Linkage Grant No: LP130100110, and supported by the Cooperative Research Centre for Advanced Automotive Technologies (AutoCRC).