International Journal of Speech Technology (2018) 21:533–544
https://doi.org/10.1007/s10772-018-9513-x

Improving the performance of GPLDA speaker verification using unsupervised inter-dataset variability compensation approaches

Ahilan Kanagasundaram[1]

Received: 26 September 2017 / Accepted: 12 April 2018 / Published online: 3 May 2018

Abstract
In practical applications, speaker verification systems have to be developed and trained using data which is outside the domain of the intended application as the collection of significant amount of in-domain data could be difficult. Experimental studies have found that when a GPLDA system is trained using out-domain data, it significantly affects the speaker verification performance due to the mismatch between development data and evaluation data. This paper proposes several unsupervised inter-dataset variability compensation approaches for the purpose of improving the performance of GPLDA systems trained using out-domain data. We show that when GPLDA is trained using out-domain data, we can improve the performance by as much as 39% by using by score normalisation using small amounts of in-domain data. Also in situations where rich out-domain data and only limited in-domain data are available, a pooled-linear-weighted technique to estimate the GPLDA parameters shows 35% relative improvements in equal error rate (EER) on int–int conditions. We also propose a novel inter-dataset covariance normalization (IDCN) approach to overcome in- and out-domain data mismatch problem. Our unsupervised IDCN-compensated GPLDA system shows 14 and 25% improvement respectively in EER over out-domain GPLDA speaker verification on tel–tel and int–int training–testing conditions. We provide intuitive explanations as to why these inter-dataset variability compensation approaches provide improvements to speaker verification accuracy.

Keywords Speaker recognition · i-Vectors · Domain adaptation · PLDA · Unsupervised approaches

1 Introduction

A significant amount of development data, especially in the presence of large intersession variability, is required to develop a speaker verification system (Vogt et al. 2008; McLaren et al. 2010; Kinnunen and Li 2010; Kanagasundaram et al. 2014a, b, 2011). During the recent Speaker and Language Recognition Summer Workshop in 2013, researchers found that when speaker verification systems are developed in data from one database and evaluated with data from another database, the dataset mismatch significantly affects the speaker verification performance (NIST 2013; Garcia-Romero and McCree 2014; Aronowitz 2014; Glembek et al. 2014). This was found to be true even in situations where both databases had the same source data—for example land-line telephone data as source.

✉ Ahilan Kanagasundaram
ahilan@eng.jfn.ac.lk

[1] Department of Electrical & Electronic Engineering, University of Jaffna, Jaffna 40000, Sri Lanka

With the current state-of-the-art speaker verification systems that have been developed, it has been found that a substantial amount of development data matched to the evaluation data is required to develop systems that achieve acceptable performance. However, in practice it is difficult to collect adequate amount of such matched data, especially speaker-labelled data in real world environments. In addressing this problem of development data/evaluation data mismatch, researchers have referred to the NIST databases as in-domain data and the Switchboard database as out-domain data (NIST 2013). These are the two main databases which are predominately used in the development of speaker verification systems. The Switchboard database was collected from a telephone switchboard (Godfrey et al. 1992). On the other hand, speech from both landline services and cellular transmissions and variety of languages are represented in the NIST database (NIST 2004, 2006, 2008, 2010).

Research effort are currently being directed to develop domain adaptation approaches to achieve state-of-the-art speaker verification performance in situations where significant amounts of labelled out-domain data are available,