Domain Mismatch Compensation for Text-Independant Speaker Recognition

Valentin Iovene
toogy@lrde.epita.fr
LRDE
Laboratoire de Recherche et Développement de l'EPITA

June 10, 2016

Abstract

Although the development of the **i-vector**-based probabilistic linear discriminant analysis (PLDA) systems led to promising results in speaker recognition, the impact of **domain mismatch** when the system training data and the evaluation data are collected from different sources remains a challenge. Johns Hopkins University (JHU) 2013 speaker recognition workshop, for which a domain adaptation challenge (DAC13) corpus was created, focused on finding solutions to address this problem.

This research report lays out the state-of-the-art techniques used for domain mismatch compensation ; such as a combination of various **whitening** transforms, and the use of a **dataset-invariant** covariance normalization to obtain domain-invariant representations of PLDA training data. Those techniques are evaluated on the DAC13 corpus and compared.

Résumé

Bien que le développement des systèmes d'analyse discriminante linéaire probabiliste (PLDA) basés sur les i-vecteurs a donné lieu à des résultats prometteurs en reconnaissance du locuteur, l'impact du *domain mismatch* lorsque les données d'entraînement du système et les données d'évaluation proviennent de sources différentes reste un défi. Le workshop de reconnaissance du locuteur de 2013 de l'Université Johns Hopkins (JHU), pour lequel un corpus d'adaptation du domaine (DAC13) a été créé, a travaillé à trouver des solutions pour résoudre ce problème.

Ce rapport de recherche présente les techniques de pointe utilisées pour la compensation du *domain mismatch* ; comme une combinaison de plusieurs transformées de blanchiment, et la normalisation de la covariance indépendante du jeu de données pour obtenir des représentations des données d'entraînement de la PLDA invariantes par rapport au domaine. Ces techniques sont évaluées sur le corpus DAC13 et comparées.

Contents

1 Domain Mismatch Compensation

This section gives a brief definition of what *domain mismatch* is and lays out examples of strategies used for compensating it. It attempts to make the context clear, ensuring the reader fully understands the next sections.

An audio recording of a speech contains various information, like:

- characteristics of the speaker's voice,
- environment specificities,
- the content of the speech: language and words used by the speaker,
- or behavioral information such as the accent or speech rate.

Depending on the task, some information can either be very valuable or completely useless (noise). It even becomes a **handicap** because the noise varying from one example to another leads to an increased error rate when comparing two speakers. Note this is not only true for the particular task of speaker recognition but also in general. Errors are observed because the decision is not based *purely* on the speaker's vocal characteristics, which is the only valuable information in the case of *text-independent* speaker recognition, but also on information that is not relevant to the task.

Domain mismatch happens when the training examples (audio recordings) used for tuning the hyperparameters [1] of the speaker recognition system and the examples used for enrolment+testing do not come from the same dataset. They were recorded in different conditions: the microphone used for recording or the environment surrounding the speaker may not have been the same.

Compensating domain mismatch means filtering out the information in the data that is specific to the domain and emphasizing the speaker-dependent information.

1.1 Whitening

1. training the universal background model and total variability matrix

2 Library of Whiteners

3 Dataset-Invariant Covariance Normalization

4 Evaluation on DAC13

4.1 About the DAC13 dataset

It is very well known.

4.2 Metrics

5 Results

Performance is shown in terms of equal error rate (EER), minimum decision cost function (minDCF) and detection error tradeoff (DET) curves.

Glossary

total variability matrix lol.

Acronyms

UBM universal background model.