論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A study on discriminative training techniques for speaker verification
著者(和文)	ﾖﾊﾝﾛﾀﾞﾝ
Author(English)	Johan Rohdin
出典(和文)	学位:博士（学術）, 学位授与機関:東京工業大学, 報告番号:甲第10021号, 授与年月日:2015年11月30日, 学位の種別:課程博士, 審査員:篠田　浩一,德永　健伸,小池　英樹,村田　剛志,藤井　敦
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10021号, Conferred date:2015/11/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（論文博士）

論　文　要　旨（英文）　（800語程度）

（Summary）

報告番号	乙 第　　　　　号	氏　名	ROHDIN Johan Andreas

（要　旨）

Chapter 1 introduces the target of speaker verification and the recent trends. It then briefly motivates the need for discriminative training (DT) and summarizes previous work on the topic. Finally, it gives a concise summary of the three contributions in this thesis which are: compensation for statistically dependent training data in DT, three constrained DT schemes and the employment of application-specific loss functions to less constrained DT schemes than previous work.
Chapter 2 introduces the basics of speaker verification. This includes terminology, evaluation metrics, as well as an overview of the standard technologies and recent trends in speaker verification.
Chapter 3 details previous work on discriminative training in speaker verification. The chapter first argues that there is a clear mismatch between the assumptions made by the model and the reality, and that therefore a discriminative training criterion is preferable. The chapter then explains discriminative training for score calibration by means of an affine transformation as well as for the PLDA model (or its score function).
The chapter ends with a discussion of three problems in current approaches to DT of PLDA-based speaker verification that will be addressed in the thesis. The first is that while GT uses utterances as observations and the speaker IDs as classes, DT uses observations from trials and *same* or *different* speaker as classes. When a training utterance (or speaker) is used in more than one trial, the trials will be statistically dependent. The second problem is that DT of PLDA easily over-fits to the training data whereas score calibration is too constrained to take full advantage of DT. The third problem is how to select loss function for a specific application.
Chapter 4 presents the experimental set-up as well as experimental results for two baselines. The first baseline is Generative PLDA training followed by score calibration by means of a discriminatively trained affine transformation of the scores. The second baseline is discriminative training of all the parameters of the PLDA score function. We refer to these two baselines as AT-Cal and Scr-UC, respectively.
Chapter 5 describes the proposal for dealing with statistically dependent training data. The chapter is divided into two parts. The first part presents a general analysis, i.e., not specific to speaker verification, of the effect of using statistically dependent training data as well as proposes how to compensate for it. To this end, we think of the training trials as random variables. The training loss for a given value of the model parameters is then also a random variable. We propose to reduce the variance of this random variable by adjusting the weight of the training trials. The method requires knowledge about the correlations between the losses of the

training trials. The second part of the chapter then proposes how to estimate these correlations for the problem at hand, discriminative training in speaker verification.

Chapter 6 proposes three new constrained DT schemes with different degree of constraints. In AT-Cal there are only two parameters to be estimated and the risk of over-fitting is therefore small. In Scr-UC on the other hand, there are more than hundred thousand parameters to be estimated and over-fitting is therefore unavoidable also with very large quantities of training data. The purpose of this chapter is to find out how much the DT should be constrained for optimal performance. The chapter then proposes the new DT schemes. The first is a transformation of the PLDA LLR score function having four parameters. The second is a scaling of each element in the i-vectors. The third is a training scheme that, like Scr-UC, updates all parameters of the PLDA LLR score function but preserves some definiteness constraints of the PLDA score function matrices that are removed by Scr-UC.

Chapter 7 describes DT with application-specific loss functions and its application to Scr-UC. In speaker verification applications. A framework for tailoring the loss function to a specific range of OPs have been previously been proposed for AT-Cal. In this chapter we are applying this framework to Scr-UC. This exploration is important because the merit of application-specific loss functions for less constrained DT schemes than AT-Cal is not obvious.

Chapter 8 presents experiments with combinations of the proposed methods.

Chapter 9 summarizes the findings in the thesis and suggests directions of future work. In conclusion, weight-adjustment is effective in particular when the number of training speakers is large or when many parameters are to be optimized by DT. Constraining DT of PLDA is important. On one of the evaluation sets, AT-Cal (the most constrained DT scheme) was the best. On the other evaluation set, one of the proposed DT schemes (the second most constrained DT scheme) was the best. We also concluded that methods for targeting as specific OP, previously proposed for AT-Cal, also works well for Scr-UC as long as an appropriate starting point for the optimization is selected.