

Discriminative Learning for Speech Recognition

Theory and Practice

Copyright © 2008 by Morgan & Claypool

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Discriminative Learning for Speech Recognition: Theory and Practice
Xiadong He and Li Deng
www.morganclaypool.com

ISBN: 9781598293081 paperback

ISBN: 9781598293098 ebook

DOI: 10.2200/S00134ED1V01Y200807SAP004

A Publication in the Morgan & Claypool Publishers series

SYNTHESIS LECTURES ON SPEECH AND AUDIO PROCESSING #4

Lecture #4

Series Editor: B. H. Juang, Georgia Institute of Technology

Series ISSN

ISSN XXX-XXXX print

ISSN XXX-XXXX electronic

Discriminative Learning for Speech Recognition

Theory and Practice

Xiaodong He and Li Deng

Microsoft Research

SYNTHESIS LECTURES ON SPEECH AND AUDIO PROCESSING #4



MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

In this book, we introduce the background and mainstream methods of probabilistic modeling and discriminative parameter optimization for speech recognition. The specific models treated in depth include the widely used exponential-family distributions and the hidden Markov model. A detailed study is presented on unifying the common objective functions for discriminative learning in speech recognition, namely maximum mutual information (MMI), minimum classification error, and minimum phone/word error. The unification is presented, with rigorous mathematical analysis, in a common rational-function form. This common form enables the use of the growth transformation (or extended Baum–Welch) optimization framework in discriminative learning of model parameters. In addition to all the necessary introduction of the background and tutorial material on the subject, we also included technical details on the derivation of the parameter optimization formulas for exponential-family distributions, discrete hidden Markov models (HMMs), and continuous-density HMMs in discriminative learning. Selected experimental results obtained by the authors in firsthand are presented to show that discriminative learning can lead to superior speech recognition performance over conventional parameter learning. Details on major algorithmic implementation issues with practical significance are provided to enable the practitioners to directly reduce the theory in the earlier part of the book into engineering practice.

KEYWORDS

Speech recognition, discriminative learning, optimization, growth transformation, hidden Markov model, exponential-family distribution

Contents

1.	Introduction and Background.....	1
1.1	What is Discriminative Learning?	1
1.2	What is Speech Recognition?.....	2
1.3	Roles of Discriminative Learning in Speech Recognition.....	4
1.4	Background: Basic Probability Distributions.....	5
1.4.1	Multinomial Distribution.....	6
1.4.2	Gaussian and Mixture-of-Gaussian Distributions.....	7
1.4.3	Exponential-Family Distribution	7
1.5	Background: Basic Optimization Concepts and Techniques.....	17
1.5.1	Basic Definitions	18
1.5.2	Necessary and Sufficient Conditions for an Optimum.....	18
1.5.3	Lagrange Multiplier Method for Constrained Optimization.....	19
1.5.4	Gradient Descent Method	20
1.5.5	Growth Transformation Method: Introduction.....	21
1.6	Organization of the Book.....	23
2.	Statistical Speech Recognition: A Tutorial	25
2.1	Introduction.....	25
2.2	Language Modeling	26
2.3	Acoustic Modeling and HMMs.....	27
3.	Discriminative Learning: A Unified Objective Function	31
3.1	Introduction.....	31
3.2	A Unified Discriminative Training Criterion	32
3.2.1	Notations	32
3.2.2	The Central Result	32
3.3	MMI and its Unified Form	33
3.3.1	Introduction to MMI Criterion	33
3.3.2	Reformulation of the MMI Criterion into Its Unified Form	34

3.4	MCE and its Unified Form	35
3.4.1	Introduction to the MCE Criterion	35
3.4.2	Reformulation of the MCE Criterion Into its Unified Form	38
3.5	Minimum Phone/Word Error and its Unified Form	39
3.5.1	Introduction to the MPE/MWE Criterion	39
3.5.2	Reformulation of the MPE/MWE Criterion Into Its Unified Form	40
3.6	Discussions and Comparisons	41
3.6.1	Discussion and Elaboration on the Unified Form	41
3.6.2	Comparisons With Another Unifying Framework	43
4.	Discriminative Learning Algorithm for Exponential-Family Distributions	47
4.1	Exponential-Family Models for Classification	47
4.2	Construction of Auxiliary Functions	48
4.3	GT Learning for Exponential-Family Distributions	49
4.4	Estimation Formulas for Two Exponential-Family Distributions	54
4.4.1	Multinomial Distribution	54
4.4.2	Multivariate Gaussian Distribution	55
5.	Discriminative Learning Algorithm for HMM	59
5.1	Estimation Formulas for Discrete HMM	59
5.1.1	Constructing Auxiliary Function $F(\Lambda; \Lambda')$	60
5.1.2	Constructing Auxiliary Function $V(\Lambda; \Lambda')$	60
5.1.3	Simplifying Auxiliary Function $V(\Lambda; \Lambda')$	61
5.1.4	GT by Optimizing Auxiliary Function $U(\Lambda; \Lambda')$	65
5.2	Estimation Formulas for CDHMM	67
5.3	Relationship with Gradient-Based Methods	70
5.4	Setting Constant D for GT-Based Optimization	71
5.4.1	Existence Proof of Finite D in GT Updates for CDHMM	72
6.	Practical Implementation of Discriminative Learning	75
6.1	Computing $\Delta\gamma(i, r, t)$ in Growth-Transform Formulas	75
6.1.1	Product Form of $C(s)$ (for MMI)	76
6.1.2	Summation Form of $C(s)$ (MCE and MPE/MWE)	78
6.2	Computing $\Delta\gamma(i, r, t)$ Using Lattices	79
6.2.1	Computing $\Delta\gamma(i, r, t)$ for MMI Involving Lattices	80
6.2.2	Computing $\Delta\gamma(i, r, t)$ for MPE/MWE Involving Lattices	83
6.2.3	Computing $\Delta\gamma(i, r, t)$ for MCE Involving Lattices	87

6.3	Arbitrary Exponent Scaling in MCE Implementation.....	88
6.4	Arbitrary Slope in Defining MCE Cost Function	89
7.	Selected Experimental Results	91
7.1	Experimental Results on Small ASR Tasks TIDIGITS.....	91
7.2	Telephony LVAR Applications.....	92
8.	Epilogue.....	97
8.1	Summary of Book Contents	97
8.2	Summary of Contributions.....	98
8.3	Remaining Theoretical Issue and Future Direction.....	99
	Major Symbols Used in the Book and Their Descriptions	103
	Mathematical Notation	105
	Bibliography.....	107
	Author Biography	111

