

Intelligent Signal Processing



Edited by
SIMON HAYKIN • BART KOSKO

INTELLIGENT SIGNAL PROCESSING

Edited by

Simon Haykin

*McMaster University
Hamilton, Ontario, Canada*

Bart Kosko

*University of Southern California
Los Angeles, CA, USA*

A Selected Reprint Volume



**IEEE
PRESS**

The Institute of Electrical and Electronics Engineers, Inc., New York

This book and other books may be purchased at a discount from the publisher when ordered in bulk quantities. Contact:

IEEE Press Marketing
Attn: Special Sales
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Fax: +1 732 981 9334

For more information about IEEE Press products, visit the IEEE Online Catalog and Store: <http://www.ieee.org/store>.

© 2001 by the Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, 17th Floor, New York, NY 10016-5997.

All rights reserved. No part of this book may be reproduced in any form, nor may it be stored in a retrieval system or transmitted in any form, without written permission from the publisher.

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

ISBN 0-7803-6010-9

IEEE Order No. PC5860

Library of Congress Cataloging-in-Publication Data

Intelligent signal processing / edited by Simon Haykin, Bart Kosko.

p. cm.

"A selected reprint volume."

Includes index.

ISBN 0-7803-6010-9

1. Signal processing—Digital techniques. 2. Intelligent control systems. 3. Adaptive signal processing. I. Haykin, Simon S., 1931– II. Kosko, Bart.

TK5102.9.I5455 2000

621.382'2—dc21

00-061369

INTELLIGENT SIGNAL PROCESSING



7911096

This book is dedicated to Bernard Widrow
for laying the foundations of adaptive filters



7711090

Preface

This expanded reprint volume is the first book devoted to the new field of intelligent signal processing (ISP). It grew out of the November 1998 ISP special issue of the *IEEE Proceedings* that the two of us coedited. This book contains new ISP material and a fuller treatment of the articles that appeared in the ISP special issue.

WHAT IS ISP?

ISP uses learning and other “smart” techniques to extract as much information as possible from incoming signal and noise data. It makes few if any assumptions about the statistical structure of signals and their environment. ISP seeks to let the data set tell its own story rather than to impose a story on the data in the form of a simple mathematical model.

Classical signal processing has largely worked with mathematical models that are linear, local, stationary, and Gaussian. These assumptions stem from the precomputer age. They have always favored closed-form tractability over real-world accuracy, and they are no less extreme because they are so familiar.

But real systems are nonlinear except for a vanishingly small set of linear systems. Almost all bell-curve probability densities have infinite variance and infinite higher order moments. The set of bell-curve densities itself is a vanishingly small set in the space of all probability densities. Real-world systems are often highly nonlinear and can depend on many partially correlated variables. The systems can have an erratic or impulsive statistical structure that varies in time in equally erratic ways. Small changes in the signal or noise structure can lead to qualitative global changes in how the system filters noise or maintains stability.

ISP has emerged recently in signal processing in much the same way that intelligent control has emerged from standard linear control theory. Researchers have guessed less at equations to model a complex system’s throughput and have instead let so-called “intelligent” or “model-free” techniques guess more for them.

Adaptive neural networks have been the most popular black box tools in ISP. Multilayer perceptrons and radial-basis function networks extend adaptive linear combiners to the nonlinear domain but require vastly more computation. Other ISP techniques include fuzzy rule-based systems, genetic algorithms, and the symbolic expert systems of artificial intelligence. Both neural and fuzzy systems can learn with supervised and unsupervised techniques. Both are (like polynomials) universal function approximators: They can uniformly approximate any continuous function on a compact domain, but this may not be practical in many real-world cases. The property of universal approximation justifies the term “model free” to describe neural and fuzzy systems even though equations describe their own throughput structure. They are one-size-fits-all approximators that can model any process if they have access to enough training data.

But ISP tools face new problems when we apply them to more real-world problems that are nonlinear, nonlocal, nonstationary, non-Gaussian, and of high dimension. Practical neural systems may require prohibitive computation to tune the values of their synaptic weights for large sets of high-dimensional data. New signal data may require total retraining or may force the neural network’s vast and unfathomable set of synapses to forget some of the signal structure it has learned. Blind fuzzy approximators need a number of if-then rules that grows exponentially with the dimension of the training data. This volume explores how the ISP tools can address these problems.

ORGANIZATION OF THE VOLUME

The 15 chapters in this book give a representative sample of current research in ISP and each has helped extend the ISP frontier. Each chapter passed through a full peer-review filter:

1. Steve Mann describes a novel technique that lets one include human intelligence in the operation of a wearable computer.
2. Sanya Mitaim and Bart Kosko present the noise processing technique of stochastic resonance in a signal processing framework and then show how neural or fuzzy or other model-free systems can adaptively add many types of noise to nonlinear dynamical systems to improve their signal-to-noise ratios.
3. Malik Magdon-Ismael, Alexander Nicholson, and Yaser S. Abu-Mostafa explore how additive noise affects information processing in problems of financial engineering.
4. Partha Niyogi, Fredrico Girosi, and Tomaso Poggio show how prior knowledge and virtual sampling can expand the size of a data set that trains a generalized supervised learning system.
5. Kenneth Rose reviews how the search technique of deterministic annealing can optimize the design of unsupervised and supervised learning systems.
6. Jose C. Principe, Ludong Wang, and Mark A. Motter use the neural self-organizing map as a tool for the local modeling of a nonlinear dynamical system.
7. Lee A. Feldkamp and Gintaras V. Puskorius describe how time-lagged recurrent neural networks can perform difficult tasks of nonlinear signal processing.
8. Davide Mattera, Francesco Palmieri, and Simon Haykin describe a semiparametric form of support vector machine for nonlinear model estimation that uses prior knowledge that comes from a rough parametric model of the system under study.
9. Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner review ways that gradient-descent learning can train a multilayer perceptron for handwritten character recognition.
10. Shigeru Katagiri, Biing-Hwang Juang, and Chin-Hui Lee show how to use the new technique of generalized probabilistic gradients to solve problems in pattern recognition.
11. Lee A. Feldkamp, Timothy M. Feldkamp, and Danil V. Prokhorov present an adaptive classification scheme that combines both supervised and unsupervised learning.
12. J. Scott Goldstein, J.R. Guescin and I.S. Reed describe an algebraic procedure based on reduced rank modeling as a basis for intelligent signal processing
13. Simon Haykin and David J. Thomson discuss an adaptive procedure for the difficult task of detecting a nonstationary target signal in a nonstationary background with unknown statistics.
14. Robert D. Dony and Simon Haykin describe an image segmentation system based on a mixture of principal components.
15. Aapo Hyvärinen, Patrik Hoyer and Erkki Oja discuss how sparse coding can denoise images.

These chapters show how adaptive systems can solve a wide range of difficult tasks in signal processing that arise in highly diverse fields. They are a humble but important first step on the road to truly intelligent signal processing.

Simon Haykin
McMaster University
Hamilton, Ontario, Canada

Bart Kosko
University of Southern California
Los Angeles, CA

List of Contributors

Chapter 1

Steve Mann
University of Toronto
Department of Electrical and Computer Engineering
10 King's College Road, S.F. 2001
Toronto, Ontario, M5S 3G4 CANADA

Chapter 2

Bart Kosko
Signal and Image Processing Institute
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, California 90089-2564

Sanya Mitaim
Signal and Image Processing Institute
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, California 90089-2564

Chapter 3

Malik Magdon-Ismael
Department of Electrical Engineering
California Institute of Technology
136-93 Pasadena, CA 91125

Alexander Nicholson
Department of Electrical Engineering
California Institute of Technology
136-93 Pasadena, CA 91125

Yaser S. Abu-Mostafa
Department of Electrical Engineering
California Institute of Technology
136-93 Pasadena, CA 91125

Chapter 4

Partha Niyogi
Massachusetts Institute of Technology
Center for Biological and Computational Learning
Cambridge, MA 02129

Fredrico Girosi
Massachusetts Institute of Technology
Center for Biological and Computational Learning
Cambridge, MA 02129

Tomaso Poggio
Massachusetts Institute of Technology
Center for Biological and Computational Learning
Cambridge, MA 02129

Chapter 5

Kenneth Rose
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Chapter 6

Jose C. Principe
Computational NeuroEngineering Laboratory
University of Florida
Gainesville, FL 32611

Ludong Wang
Computational NeuroEngineering Laboratory
University of Florida
Gainesville, FL 32611

Mark A. Motter
Computational NeuroEngineering Laboratory
University of Florida
Gainesville, FL 32611

Chapter 7

Lee A. Feldkamp
Ford Research Laboratory
PO Box 2053
Dearborn, MI 48121-2053

Gintaras V. Puskorius
Ford Research Laboratory
PO Box 2053
Dearborn, MI 48121-2053

Chapter 8

Davide Mattera
Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni
Università degli Studi di Napoli Federico II
Via Claudio, 21, 80125
Napoli, ITALY

Francesco Palmieri
Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni
Università degli Studi di Napoli Federico II
Via Claudio, 21, 80125
Napoli, ITALY

Simon Haykin
McMaster University
Electrical and Computer Engineering
1280 Main Street West
Hamilton, Ontario, L8S 4K1 CANADA

Chapter 9

Yann LeCun
Speech and Image Processing Services Research Laboratory
AT&T Labs—Research
100 Schulz Drive
Red Bank, NJ 07701

Léon Bottou
Speech and Image Processing Services Research Laboratory
AT&T Labs—Research
100 Schulz Drive
Red Bank, NJ 07701

Yoshua Bengio
Speech and Image Processing Services Research Laboratory
AT&T Labs—Research
100 Schulz Drive
Red Bank, NJ 07701
and

Département d'Informatique et de Recherche Opérationnelle
Université de Montréal
C.P. 6128
Succ. Centre-Ville
2920 Chemin de la Tour
Montréal, Québec, H3C 3J7 CANADA

Patrick Haffner
Speech and Image Processing Services Research Laboratory
AT&T Labs—Research
100 Schulz Drive
Red Bank, NJ 07701

Chapter 10

Shigeru Katagiri
ATR Human Information Processing Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-02 JAPAN

Biing-Hwang Juang
Bell Laboratories, Lucent Technologies
600–700 Mountain Avenue
Murray Hill, NJ 07974-0636

Chin-Hui Lee
Bell Laboratories, Lucent Technologies
600–700 Mountain Avenue
Murray Hill, NJ 07974-0636

Chapter 11

Lee A. Feldkamp
Powertrain Control Systems Department
Ford Research Laboratory
Ford Motor Company
Dearborn, MI 48121-2053

Timothy M. Feldkamp
Powertrain Control Systems Department
Ford Research Laboratory
Ford Motor Company
Dearborn MI 48121-2053

Danil V. Prokhorov
Powertrain Control Systems Department
Ford Research Laboratory
Ford Motor Company
Dearborn, MI 48121-2053

Chapter 12

J. Scott Goldstein
SAIC
Adaptive Signal Exploitation
4001 Fairfax Drive, Suite 675
Arlington, VA 22203

List of Contributors

J. R. Guerci
SAIC
Adaptive Signal Exploitation
4001 Fairfax Drive, Suite 675
Arlington, VA 22203

I. S. Reed
Department of Electrical Engineering—Systems
University of Southern California
Los Angeles, CA 90089-2564

Chapter 13

Simon Haykin
McMaster University
Electrical and Computer Engineering
1280 Main Street West
Hamilton, Ontario, L8S 4K1 CANADA

David J. Thomson
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974

Chapter 14

Robert D. Dony
School of Engineering
University of Guelph
Guelph, ON N1G 2W1 CANADA

Simon Haykin
McMaster University
Electrical and Computer Engineering
1280 Main Street West
Hamilton, Ontario, L8S 4K1 CANADA

Chapter 15

Aapo Hyvärinen
Helsinki University of Technology
Laboratory of Computer and Information Science
P.O.B. 5400
Espoo 02015 FINLAND

Patrik Hoyer
Neural Networks Research Centre
Helsinki University of Technology
PO Box 5400, FIN-02015 HUT
FINLAND

Erkki Oja
Helsinki University of Technology
Rakentajanaukio 2C
Espoo 02015 FINLAND

Contents

Preface xvii

List of Contributors xix

CHAPTER 1	Humanistic Intelligence: “Wear Comp” As a New Framework and Application for Intelligent Signal Processing	1
	INTRODUCTION	1
	Why Humanistic Intelligence?	2
	Humanistic Intelligence Does Not Necessarily Mean “user-friendly”	2
	“WEAR COMP” AS MEANS OF REALIZING HUMANISTIC INTELLIGENCE	3
	Basic Principals of WearComp	3
	<i>Operational modes of WearComp</i>	3
	The Six Basic Signal Flow Paths of WearComp	5
	PHILOSOPHICAL ISSUES	6
	Fundamental Issues of WearComp	6
	<i>Historical context</i>	6
	<i>The shift from guns to cameras and computers</i>	6
	<i>The shift from draconian punishment to micro management</i>	7
	<i>Fundamental issues of WearComp</i>	8
	Aspects of WearComp and Personal Empowerment	9
	PRACTICAL EMBODIMENT OF WEARCOMP	10
	Building Signal-Processing Devices Directly into Fabric	12
	<i>Remaining issue of underwearable signal processing hardware</i>	13
	Multidimensional Signal Input for Humanistic Intelligence	14
	<i>Safety first</i>	15
	<i>More than just a health status monitor</i>	16
	PERSONAL IMAGING APPLICATION OF HUMANISTIC INTELLIGENCE	16
	Some Simple Illustrative Examples	16
	<i>Always ready: From point and click to “look and think”</i>	16
	<i>Personal safety device for reducing crime</i>	17
	<i>The retro-autofocus example: Human in the signal processing loop</i>	17
	Mathematical Framework for Personal Imaging	18
	<i>Quantigraphic imaging and the Wyckoff Principle</i>	18
	<i>Video orbits</i>	20

<i>Dynamic range and “dynamic domain”</i>	24
<i>Bi-foveated WearCam</i>	30
<i>Lightspace modeling for HI</i>	31
BEYOND VIDEO: SYNTHETIC SYNESTHESIA AND PERSONAL IMAGING	34
Synthetic Synesthesia: Adding New Sensory Capabilities to the Body	34
<i>Safety first</i>	35
<i>A true extension of the mind and body</i>	36
CONCLUSIONS	36
ACKNOWLEDGMENTS	37
REFERENCES	37

CHAPTER 2

Adaptive Stochastic Resonance	40
ABSTRACT	40
STOCHASTIC RESONANCE AND ADAPTIVE FUNCTION APPROXIMATION	41
SR DYNAMICAL SYSTEMS	46
SR PERFORMANCE MEASURES	51
ADDITIVE FUZZY SYSTEMS AND FUNCTION APPROXIMATION	56
SR LEARNING AND EQUILIBRIUM	58
The Signal-to-Noise Ratio in Nonlinear Systems	58
Supervised Gradient Learning and SR Optimality	63
SR LEARNING: SIMULATION RESULTS	68
SR Test Case: The Quartic Bistable System	70
Other SR Test Cases	77
Fuzzy Sr Learning: The Quartic Bistable System	81
CONCLUSIONS	86
REFERENCES	89
APPENDIX A. THE STANDARD ADDITIVE MODEL (SAM) THEOREM	104
APPENDIX B. SAM GRADIENT LEARNING	106

CHAPTER 3

Learning in the Presence of Noise	108
ABSTRACT	108
INTRODUCTION	108
FINANCIAL TIME SERIES PREDICTION	110
IMPACT OF NOISE ON LEARNING	111
The Learning Problem	111
Performance of a Learning System	113
Estimating the Model Limitation	116
APPLICATION TO FINANCIAL MARKET FORECASTING	116
CONCLUSION	117
ACKNOWLEDGMENTS	118
APPENDIX	119
REFERENCES	125

CHAPTER 4

Incorporating Prior Information in Machine Learning by Creating Virtual Examples	127
ABSTRACT	127
LEARNING FROM EXAMPLES	128
Background: Learning as Function Approximation	128
PRIOR INFORMATION AND THE PROBLEM OF SAMPLE COMPLEXITY	129

VIRTUAL EXAMPLES: A FRAMEWORK FOR PRIOR INFORMATION	133
The General Framework	134
Techniques for Prior Information and Related Research	135
<i>Prior Knowledge in the Choice of Variables of Features</i>	135
<i>Prior Knowledge in the Learning Technique</i>	135
<i>Generating New Examples with Prior Knowledge</i>	136
<i>Incorporating Prior Knowledge as Hints</i>	137
VIRTUAL EXAMPLES AND REGULARIZATION	138
Regularization Theory and RBF	138
Regularization Theory in Presence of Radial Symmetry	140
Radial Symmetry and "Virtual" Examples	141
VIRTUAL EXAMPLES IN VISION AND SPEECH	143
Virtual Views for Object Recognition	144
Symmetry as Prior Information	146
More General Transformations: Linear Object Classes	147
<i>3D Objects, 2D Projections, and Linear Classes</i>	148
<i>Implications</i>	149
<i>Learning the Transformation</i>	150
Virtual Examples in Speech Recognition	152
CONCLUSIONS	158
REFERENCES	158

CHAPTER 5

Deterministic Annealing for Clustering, Compression, Classification, Regression, and Speech Recognition 163

ABSTRACT	163
INTRODUCTION	163
DETERMINISTIC ANNEALING FOR UNSUPERVISED LEARNING	166
Clustering	166
<i>Principled Derivation of Deterministic Annealing</i>	167
<i>Statistical Physics Analogy</i>	169
<i>Mass-Constrained Clustering</i>	172
<i>Preferred Implementation of the DA Clustering Algorithm</i>	177
<i>Illustrative Examples</i>	178
Extensions and Applications	180
<i>Vector Quantization for Noisy Channels</i>	180
<i>Entropy-Constrained Vector Quantizer Design</i>	182
<i>Structurally Constrained Vector Quantizer Design</i>	183
<i>Graph-Theoretic and Other Optimization Problems</i>	185
DETERMINISTIC ANNEALING FOR SUPERVISED LEARNING	188
Problem Formulation	188
Basic Derivation	190
Generality and Wide Applicability of the DA Solution	192
<i>Regression, Classification, and Clustering</i>	192
<i>Structures</i>	193
Experimental Results	197
<i>VQ Classifier Design</i>	197
<i>RBF Classifier Design</i>	201
<i>MLP Classifier Design</i>	201
<i>Piecewise Regression</i>	202
<i>Mixture of Experts</i>	206

SPEECH RECOGNITION	207
Problem Formulation	212
Design by Deterministic Annealing	213
Experimental Results	215
THE RATE-DISTORTION CONNECTION	217
RECENT DA EXTENSIONS	219
SUMMARY	220
REFERENCES	221

CHAPTER 6 Local Dynamic Modeling with Self-Organizing Maps and Applications to Nonlinear System Identification and Control 230

ABSTRACT	230
INTRODUCTION	230
DYNAMIC MODELING	232
GLOBAL AND LOCAL MODELS	235
Global Dynamic Models	236
Local Dynamic Models	237
State Dependent Prediction of Nonlinear AR Processes	239
KOHONEN'S SELF-ORGANIZING MAP (SOM)	240
SOM Networks and Kohonen Learning	240
Codebook in Reconstruction Space	243
SOM-BASED LOCAL LINEAR MODELS	245
Dynamic Learning in the SOM	248
EXPERIMENTAL RESULTS	251
DESIGN OF SWITCHING CONTROLLERS BASED ON SOM	255
Problem Specification	255
The Control Strategy	255
Design of the PPMSC	258
EXPERIMENTAL CONTROL OF THE WIND TUNNEL	260
Comparison of PPMSC Control to Existing Automatic Controller and Expert Operator	260
CONCLUSIONS	264
REFERENCES	267
APPENDIX	270

CHAPTER 7 A Signal Processing Framework Based on Dynamic Neural Networks with Application to Problems in Adaptation, Filtering and Classification 272

ABSTRACT	272
INTRODUCTION	272
NETWORK ARCHITECTURE AND EXECUTION	273
GRADIENT CALCULATION	275
EKF MULTI-STREAM TRAINING	275
The Kalman Recursion	275
Multi-Stream Training	277
Some Insight into the Multi-Stream Technique	278
Advantages and Extensions of Multi-Stream Training	279
SYNTHETIC EXAMPLES	280
Multiple Series Prediction and Dropout Rejection	280
Training	280

<i>Testing</i>	280
<i>Discussion</i>	281
Modeling with Stability Enforcement	282
<i>Training without Constraints</i>	283
<i>Training with Constraints</i>	284
<i>Discussion</i>	284
AUTOMOTIVE EXAMPLES	285
Sensor-Catalyst Modeling	285
<i>Experimental Data</i>	287
<i>Training and Testing</i>	287
Engine Misfire Detection	289
<i>Applying Recurrent Networks</i>	290
SUMMARY AND CONCLUSIONS	292
ACKNOWLEDGMENTS	292
REFERENCES	292

CHAPTER 8 **Semiparametric Support Vector Machines for Nonlinear Model Estimation** 295

ABSTRACT	295
INTRODUCTION	295
PROBLEM SETTING	296
OPTIMUM LINE SEARCH FOR A GENERIC DIRECTION	297
SIMPLE ALGORITHMS FOR PURELY NONPARAMETRIC APPROACH	298
A Gradient-Based Approach	299
A Coordinate Descent Approach	299
THE CLASSIC SVM PROBLEM	300
GENERAL SEMIPARAMETRIC SVM	301
The Augmented Lagrangian Approach	301
Training Semiparametric SVM	301
<i>A Possible Simplification</i>	302
REGULARIZED PARAMETRIC SVM APPROACH	302
EXPERIMENTAL RESULTS	302
CONCLUSIONS	305
ACKNOWLEDGMENTS	305
REFERENCES	305

CHAPTER 9 **Gradient-Based Learning Applied to Document Recognition** 306

INTRODUCTION	306
Learning from Data	307
Gradient-Based Learning	308
Gradient Back-Propogation	308
Learning in Real Handwriting Recognition Systems	309
Globally Trainable Systems	309
CONVOLUTIONAL NEURAL NETWORKS FOR ISOLATED CHARACTER RECOGNITION	310
Convolutional Networks	311
LeNet-5	312
Loss Function	314

RESULTS AND COMPARISON WITH OTHER METHODS	314
Database: the Modified NIST Set	314
Results	315
Comparison with Other Classifiers	316
<i>Linear Classifier, and Pairwise Linear Classifier</i>	316
<i>Baseline Nearest Neighbor Classifier</i>	317
<i>Principal Component Analysis (PCA) and Polynomial Classifier</i>	317
<i>Radial Basis Function Network</i>	318
<i>One-Hidden Layer Fully Connected Multilayer Neural Network</i>	318
<i>Two-Hidden Layer Fully Connected, Multilayer Neural Network</i>	318
<i>A Small Convolutional Network: LeNet-1</i>	318
<i>LeNet-4</i>	318
<i>Boosted LeNet-4</i>	318
<i>Tangent Distance Classifier (TDC)</i>	319
<i>Support Vector Machine (SVM)</i>	319
Discussion	319
Invariance and Noise Resistance	320
MULTI-MODULE SYSTEMS AND GRAPH TRANSFORMER NETWORKS	321
An Object-Oriented Approach	321
Special Modules	322
Graph Transformer Networks	323
MULTIPLE OBJECT RECOGNITION: HEURISTIC OVER-SEGMENTATION	324
Segmentation Graph	324
Recognition Transformer and Viterbi Transformer	324
GLOBAL TRAINING FOR GRAPH TRANSFORMER NETWORKS	326
Viterbi Training	326
Discriminative Viterbi Training	327
Forward Scoring, and Forward Training	329
Discriminative Forward Training	330
Remarks on Discriminative Training	331
MULTIPLE OBJECT RECOGNITION: SPACE DISPLACEMENT NEURAL NETWORK	332
Interpreting the Output of an SDNN with a GTN	333
Experiments with SDNN	333
Global Training of SDNN	334
Object Detection and Spotting with SDNN	335
GRAPH TRANSFORMER NETWORKS AND TRANSDUCERS	336
Previous Work	336
Standard Transduction	336
Generalized Transduction	336
Notes on the Graph Structures	338
GTN and Hidden Markov Models	339
AN ON-LINE HANDWRITING RECOGNITION SYSTEM	339
Preprocessing	339
Network Architecture	340
Network Training	341
Experimental Results	341
A CHECK READING SYSTEM	342
A GTN for Check Amount Recognition	342
Gradient-Based Learning	344

Rejecting Low Confidence Checks	344
Results	344
CONCLUSIONS	345
APPENDICES	345
Pre-conditions for Faster Convergence	345
Stochastic Gradient vs. Batch Gradient	346
Stochastic Diagonal Levenberg-Marquardt	346
ACKNOWLEDGMENTS	348
REFERENCES	348

CHAPTER 10

Pattern Recognition Using A Family of Design Algorithms Based Upon Generalized Probabilistic Descent Method 352

ABSTRACT	352
INTRODUCTION	353
Speech Pattern Recognition Using Modular Systems	354
Classifier Design Based on Maximum-Likelihood Method	357
Classifier Design Based on Discriminant Function Approach	358
Motivation of GPD Development and Paper Organization	365
DISCRIMINATIVE PATTERN CLASSIFICATION	366
Bayes Decision Theory	366
Minimum Error Rate Classification	368
Discriminant Function Approach	370
Probabilistic Descent Method	375
<i>Formalization</i>	375
<i>Probabilistic descent theorem</i>	378
<i>Problems</i>	381
GENERALIZED PROBABILISTIC DESCENT METHOD	386
Formulation Concept	386
GPD Embodiment for Distance Classifier	386
<i>Discriminant function for dynamic patterns</i>	387
<i>Formulation</i>	387
Design Optimality	390
Minimum Classification Error Learning	392
Speech Recognition Using MCE/GPD-trained Distance Classifiers	396
Remarks	399
RELATIONSHIPS BETWEEN MCE/GPD AND OTHERS	400
Relation with LVQ	400
Relation with Maximization of Mutual Information	402
Relation with Minimization of Squared Error Loss	403
DERIVATIVES OF GPD	405
Overview	405
Segmental-GPD for Continuous Speech Recognition	406
Minimum Error Training for Open-Vocabulary Recognition	410
<i>Open-vocabulary speech recognition</i>	410
<i>Minimum spotting error learning</i>	411
<i>Discriminative utterance verification</i>	415
Discriminative Feature Extraction	417
<i>Fundamentals</i>	417
<i>An example implementation for spectrum-based speech recognition</i>	420