



# PATTERN RECOGNITION

STATISTICAL, STRUCTURAL  
AND NEURAL APPROACHES

ROBERT SCHALKOFF

# *Pattern Recognition: Statistical, Structural and Neural Approaches*

**Robert J. Schalkoff**

Clemson University



John Wiley & Sons, Inc.

New York • Chichester • Brisbane • Toronto • Singapore

# Contents

---

---

<b>PART 1 INTRODUCTION AND GENERAL PATTERN RECOGNITION CONCERNS</b>	1
<b>CHAPTER 1 PATTERN RECOGNITION (PR) OVERVIEW</b>	2
<b>What is This Book About?</b>	2
Overview	2
‘Engineering’ Approaches to Pattern Recognition	3
Relationship of PR to Other Areas	3
Pattern Recognition Applications	3
<b>Pattern Recognition, Classification, and Description</b>	4
Abstract Representation of Pattern Mappings	4
Structure of a ‘Typical’ PR System	5
<b>Patterns and Feature Extraction with Examples</b>	6
Patterns and Features	6
Pattern Distortions—A Fundamental and Difficult Problem	8
Example: Feature Extraction Using Generalized Cylinders for 3-D Object Description and Classification	8
Example: Generating RST Invariant Features and Application to 2-D Figure Recognition	10
Feature Extraction	11
Numerical Results and Analysis	12
Analysis	12
What’s Really Different Among Different Patterns?	13
The Feature Vector and Feature Space	13
Definitions	14

Classifiers, Decision Regions and Boundaries, and Discriminant Functions	15
<b>Training and Learning in PR Systems</b>	16
Using A Priori Knowledge or ‘Experience’	16
Learning Curves (Improved ‘Performance’)	17
Training Approaches	17
<b>Pattern Recognition Approaches</b>	18
The Statistical Pattern Recognition Approach (StatPR)	18
The Syntactic Pattern Recognition Approach (SyntPR)	19
The Neural Pattern Recognition Approach (NeurPR)	20
Comparing and Relating StatPR, SyntPR and NeurPR	20
Examples of Pattern Recognition Approaches	20
StatPR	21
SyntPR	26
Example: Combinations of StatPR and SyntPR	26
Procedures for PR System Engineering	26
<b>Other Approaches to PR</b>	27
‘Black Box’ Approaches	27
Reasoning-Driven Pattern Recognition	28
<b>Overview of PR Literature and Resources</b>	28
Books	29
Journals	29
Conferences	30
<b>Exercises</b>	30
<b>PART 2 STATISTICAL PATTERN RECOGNITION (StatPR)</b>	33
<b>CHAPTER 2 INTRODUCTION TO STATISTICAL PATTERN RECOGNITION</b>	34
<b>Introduction to Statistical Pattern Recognition</b>	34
Approaches to Developing StatPR Classifiers	34
Simple Examples of Statistical Models and Applications	35
<b>The Gaussian Case and Class Dependence</b>	39
Gaussian Models for $p(\underline{x} w_i)$	39
<b>Discriminant Functions</b>	39
Generalized Results (Gaussian Case)	40
Decision Surfaces for Specific Cases	40
A General Look at the StatPR Approach Structure	43
<b>Additional Examples</b>	43
Example: Uniform Densities, $c = 2, d = 1$	43
Example: $c = 3, d = 2$ , Gaussian Case	44
<b>Extensions</b>	47
Training	47

Alternative Classification Procedures	47
Unsupervised Approaches	48
<b>Classifier Performance, Risk, and Errors</b>	49
Measures of Classification Performance	49
General Measures of Classification Risk	51
<b>Bibliographical Remarks</b>	54
<b>Exercises</b>	54
<b>CHAPTER 3 SUPERVISED LEARNING (TRAINING) USING PARAMETRIC AND NONPARAMETRIC APPROACHES</b>	58
<b>Introduction</b>	58
<b>Parametric Estimation and Supervised Learning</b>	58
Approaches to Parameter Estimation	59
Supervised versus Unsupervised Learning	59
<b>Maximum Likelihood (ML) Estimation</b>	59
Formulation	59
Use of the Training Set	59
The Likelihood Function	60
<b>The Bayesian Parameter Estimation Approach</b>	62
<b>Supervised Learning Using Nonparametric Approaches</b>	66
Nonparametric Approaches	66
General Aspects of Nonparametric Density Estimation	67
The Basic Approach	67
<b>Parzen Windows</b>	70
Unit Step Function $\phi(\underline{x})$	70
Extension to More General Interpolation Functions	73
<b>k-nn Nonparametric Estimation</b>	74
A Problem with $V_0$	74
Estimation of $P(w_i \underline{x})$ Directly	75
<b>Examples of Nonparametric Learning</b>	75
<b>Direct Classification Using the Training Set [The     Nearest Neighbor Rule (NNR)]</b>	75
The NNR Approach	75
<b>Bibliographical Remarks</b>	83
<b>Exercises</b>	83
<b>CHAPTER 4 LINEAR DISCRIMINANT FUNCTIONS AND THE DISCRETE AND BINARY FEATURE CASES</b>	89
<b>Introduction</b>	89
Linear Discriminant Functions	89
Fisher's Linear Discriminant	90
<b>Discrete and Binary Classification Problems</b>	94

Classification Procedures for Discrete Feature Data	94
Formulation for the Binary Case	95
Classification Procedures for the Binary Case	96
Nonparametric Estimation of $P(\underline{x} w_i)$	98
<b>Techniques to Directly Obtain Linear Classifiers</b>	99
The Concept of Linear Separability	99
Design of Linear Classifiers	99
<b>Bibliographical Remarks</b>	104
<b>Exercises</b>	105
<b>CHAPTER 5 UNSUPERVISED LEARNING AND CLUSTERING</b>	109
<b>Formulation of Unsupervised Learning Problems</b>	109
A Modified Training Set	109
Unsupervised Learning Approaches	110
<b>Clustering for Unsupervised Learning and Classification</b>	114
The Clustering Concept and the Search for ‘Natural Clusters’	114
The <i>c</i> -means Algorithm	115
Learning Vector Quantization (LVQ)	116
Formal Characterization of General Clustering Procedures	117
Clustering Strategies	118
‘Cluster Swapping’ Approaches	119
Extended Example: A Hierarchical Clustering Procedure	120
<b>Bibliographical Remarks</b>	121
<b>Exercises</b>	121
<b>PART 3 SYNTACTIC PATTERN RECOGNITION (SyntPR)</b>	127
<b>CHAPTER 6 OVERVIEW</b>	128
<b>Syntactic Pattern Recognition (SyntPR) Overview</b>	128
<b>Quantifying Structure in Pattern Description and Recognition</b>	129
Hierarchical Approaches	129
Relational Models	130
<b>Grammar-Based Approach and Applications</b>	130
Using Formal Linguistics to Model and Describe Patterns	130
Structural Analysis and Description Applications	131
<b>Elements of Formal Grammars</b>	133
Definitions and Conventions	133
Grammar and Languages Using Strings	135
Grammar Application Modes	136
Grammar Types and Productions	137
Other Production Constraints in String Grammars	139
Additional Introductory Concepts	141

Example: Context-Sensitive, Context-Free, and Finite State (Regular) Grammar Productions	142
<b>Examples of String Generation as Pattern Description</b>	144
Example: 2-D Line Drawing Description Grammar [Shaw 1970]	144
Example: Character Description Using PDL	146
Example: Object Description Using Projected Cylinder Models	146
Extended Example: Blocks World Description	146
Remarks on the Heuristic Generation of Grammars	149
<b>Bibliographical Remarks</b>	150
<b>Exercises</b>	150
<b>CHAPTER 7 SYNTACTIC RECOGNITION VIA PARSING AND OTHER GRAMMARS</b>	155
<b>Recognition of Syntactic Descriptions</b>	155
Recognition by String Matching	155
Recognition by Parsing	156
<b>Parsing</b>	156
Fundamental Concepts	156
An Abstract View of the Parsing Problem	157
Parsing Approaches	159
<b>The Cocke-Younger-Kasami (CYK) Parsing Algorithm</b>	160
The Approach	160
The CYK Table	160
<b>(Augmented) Transition Networks (ATNs) in Parsing</b>	163
Transition Networks	163
Augmented Transition Nets (ATNs)	164
<b>Higher Dimensional Grammars</b>	165
Introduction	165
Tree Grammars	166
Traversing and Describing Trees	166
Productions and Derivations with Trees	169
<b>Stochastic Grammars and Applications</b>	170
Stochastic Grammars	171
Derivations in a Stochastic Language	172
<b>Bibliographical Remarks</b>	173
<b>Exercises</b>	174
<b>CHAPTER 8 GRAPHICAL APPROACHES TO SyntPR</b>	176
<b>Graph-Based Structural Representations</b>	176
From Digraphs to Semantic Nets to Relational Graphs	176
Application of Relational Graphs to PR	177
<b>Graph Isomorphism</b>	179

Uniqueness of Digraph-Based Structural Description and Isomorphism	179
Determining Isomorphism	180
Extensions to the Elementary Graph Matching Approach	181
A Formal Characterization of the Relational Graph Similarity Problem	182
Extensions of Relational Graphs (Attributed Graphs)	183
<b>A Structured Strategy to Compare Attributed Graphs</b>	185
Relaxing Requirements for Pattern Isomorphism and Subisomorphism	185
Match Graphs (MGs)	185
Recursive Procedure to Find Cliques [ <i>cliques</i> ( $X, Y$ )]	186
<b>Other Attributed Graph Distance or Similarity Measures</b>	187
Design and Selection of Similarity Measures	187
Transforming $G_i$ into $G_j$	188
Extended Example: Structural Unification Using Attributed Graphs	188
<b>Bibliographical Remarks</b>	191
<b>Exercises</b>	192
<b>CHAPTER 9 LEARNING VIA GRAMMATICAL INFERENCE</b>	194
<b>Learning Grammars</b>	194
Problem Overview	194
Difficulties in Structural Learning	194
<b>Problem Formulation</b>	195
Characterizing the Grammar Source, $\Psi$	195
Concerns Related to the Training Set	195
<b>Grammatical Inference (GI) Approaches</b>	197
Grammatical Inference Objectives	197
Intuitive Sample GI Procedure	197
<b>Procedures to Generate Constrained Grammars</b>	198
Using a Single String to Partially Infer $G_{learn}$	198
Generation of Canonical Definite Grammars	199
General Procedure to Generate CDFSG, ( $G_c$ )	199
<b>Bibliographical Remarks</b>	200
<b>Exercises</b>	201
<b>PART 4 NEURAL PATTERN RECOGNITION (NeurPR)</b>	203
<b>CHAPTER 10 INTRODUCTION TO NEURAL NETWORKS</b>	204
<b>Neurons and Neural Nets</b>	204
Introduction	204



History	205
Neural Networks as a Black Box Approach and ‘Artificial’ Neural Systems	205
Key Neural Network Concepts	206
Characteristics of Neural Computing Applications	206
‘Connectionist’ Models and Computing	207
<b>Neural Network Structures for PR Applications</b>	207
Neural Network Structure (How Should I Connect My Neurons?)	207
Learning in Neural Networks	208
The Neural PR Application Design Phase	210
Why Is Neural Computation So Important?	210
<b>Physical Neural Networks</b>	211
The Physical Neuron	211
Dynamics of a Biological Neural System	212
<b>The Artificial Neural Network Model</b>	213
Artificial Neuron Activation and Output Characteristics	213
Neural Unit Interconnection Strengths (Weights)	217
Other Neural Network Parameters	218
Network Implementations for Constraint Satisfaction and Matching Problems	219
Hardware Realizations of Neural Networks	220
<b>Bibliographical Remarks</b>	220

<b>CHAPTER 11 INTRODUCTION TO NEURAL PATTERN ASSOCIATORS AND MATRIX APPROACHES</b>	221
<b>Neural Network-Based Pattern Associators</b>	221
Introduction	221
Design Procedure	221
‘Black Box’ Structure	222
CAM and Other Neural Memory Structures	223
Desirable Pattern Associator (PA) Properties	224
<b>Matrix Approaches (Linear Associative Mappings) and Examples</b>	224
An Elementary Linear Network Structure and Mathematical Representation	224
Approach 1: A Linear CAM (‘Hopfield’) Network	226
Approach 2: Matched Filters/Adaptive Filters	227
Approach 3: Outer Product Formulations	228
Approach 4: Generalized Inverse Applications	229
Extended Example: Heteroassociative Memory Design	231
Hebbian or Correlation-Based Learning	232
<b>Bibliographical Remarks</b>	233
<b>Exercises</b>	233

<b>CHAPTER 12 FEEDFORWARD NETWORKS AND TRAINING BY BACKPROPAGATION</b>	236
<b>Multilayer, Feedforward Network Structure</b>	236
Introduction	236
Overall Feedforward Structure	236
The Role of Internal/Hidden Layers (What's 'Hidden?')	238
Training Considerations	239
<b>Training the Feedforward Network: The Delta Rule (DR) and Generalized Delta Rule (GDR)</b>	240
Overview	240
Gradient Descent Approaches for Training	241
Training by Sample or Training by Epoch	241
Derivation of the DR	242
A General Assessment of Error Sensitivity	242
<b>Extension of the DR for Units in the Hidden Layers [The Generalized Delta Rule (GDR)]</b>	246
Back Propagation—Summary of the Multistep Procedure	247
Adding <i>Momentum</i> to the Training Procedure	247
Training the Unit Bias Inputs	249
Example: 'Handworked' Training of a Single Unit Using the GDR	250
<b>Extended Example: Pattern Associator for Character Classification</b>	251
Chosen Features	251
Case 1: Single 4-Valued Output, Pseudoinverse Solution	252
Case 2: Single 4-Valued Output, 2-Layer Neural Network Solution	252
Case 3: 4 Output ('One-of') 2-Layer Neural Network Solution	253
Case 3: Summary of Results	254
Case 4: 4-Output, 3-Layer (Hidden) Neural Network Solution	254
Other Ramifications of Hidden Units	257
<b>Bibliographical and Historical Remarks</b>	258
<b>Exercises</b>	259
<b>CHAPTER 13 CONTENT ADDRESSABLE MEMORY APPROACHES AND UNSUPERVISED LEARNING IN NeurPR</b>	264
<b>Introduction</b>	264
<b>The Hopfield Approach to Neural Computing</b>	265
Network Parameters	265
Network Dynamics	265
Hamming Distance and Its Significance in CAM	267
Example: Hamming Distances Between Stored States	267
Example: Design of a Simple Hopfield Network—Storing and Accessing Stable States	268
Why the Hopfield Strategy Should Work	271

---

<b>Additional Examples of CAM Applications in PR</b>	272
Example: Character Recognition	272
Example: Relational Constraint Satisfaction (Coloring)	273
Example: Symbolic Constraint Satisfaction (Labeling)	275
<b>Unsupervised Learning in NeurPR: Self-Organizing Networks</b>	275
Introduction	275
Adaptive Resonance Architectures	276
Self-Organizing Feature Maps (Kohonen)	280
<b>Summary</b>	284
<b>Bibliographical Remarks</b>	285
<b>Exercises</b>	286
<b>APPENDIX 1 Linear Algebra Review</b>	289
<b>APPENDIX 2 Probability and Random Variables/Vectors</b>	298
<b>APPENDIX 3 Discrete Mathematics Review</b>	314
<b>APPENDIX 4 Descent Procedures: Examples and Error Function</b>	
Contours	323
<b>APPENDIX 5 Similarity Measures, Matching Techniques, and</b>	
<b>Scale-Space Approaches</b>	328
<b>APPENDIX 6 Geometry for Classification and State-Space</b>	
Visualization	337
<b>REFERENCES</b>	344
<b>PERMISSION SOURCE NOTES</b>	360
<b>INDEX</b>	361