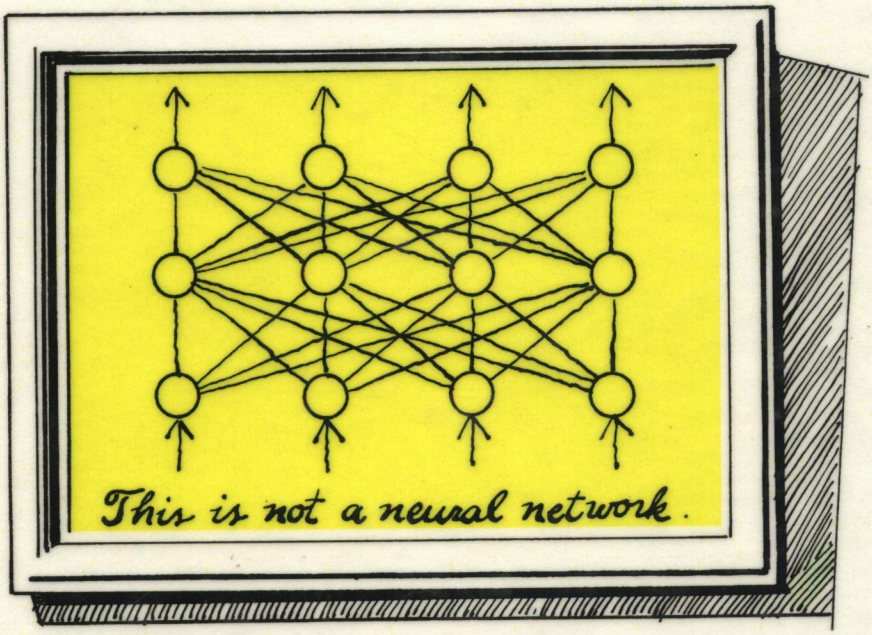
CONNECTIONIST SPEECH RECOGNITION

A Hybrid Approach

Hervé A. Bourlard
Nelson Morgan

foreword by Richard Lippmann



Kluwer Academic Publishers

Contents

List of Figures xiii

List of Tables xv

Notation xix

Foreword xxiii

Preface xxv

Acknowledgments xxvii

I BACKGROUND 1

1 INTRODUCTION 3

1.1 Automatic Speech Recognition (ASR) 4

1.2 Limitations in Current ASR Systems 9

1.3 Book Overview 10

2 STATISTICAL PATTERN CLASSIFICATION 15

2.1 Introduction 15

2.2 A Model for Pattern Classification 16

2.3 Statistical Classification 18

2.4 Pattern Classification with Realistic Data 23

2.5 Summary 25

3 HIDDEN MARKOV MODELS 27

3.1 Introduction 27

3.2 Definition and Underlying Hypotheses 31

3.3 Parametrization and Estimation 33
3.3.1 General Formulation 33
3.3.2 Continuous Input Features 38
3.3.3 Discrete Input Features 39
3.3.4 Maximum Likelihood Criterion 40
3.3.5 Viterbi Criterion 43
3.4 Training Problem 44
3.4.1 Maximum Likelihood Criterion 45
3.4.2 Viterbi Criterion 47
3.5 Decoding Problem 50
3.5.1 Maximum Likelihood Criterion 50
3.5.2 Viterbi Criterion 51
3.6 Likelihood and Discrimination 52
3.7 Summary 56

4 MULTILAYER PERCEPTRONS 59
4.1 Introduction 59
4.2 Linear Perceptrons 60
4.2.1 Linear Discriminant Function 60
4.2.2 Least Mean Square Criterion 61
4.2.3 Normal Density 62
4.3 Multilayer Perceptrons (MLP) 63
4.3.1 Some History 63
4.3.2 Motivations 64
4.3.3 Architecture and Training Procedure 66
4.3.4 Lagrange Multipliers 68
4.3.5 Speeding up EBP 71
4.3.6 On-Line and Off-Line Training 73
4.4 Nonlinear Discrimination 74
4.4.1 Nonlinear Functions in MLPs 74
4.4.2 Phonemic Strings to Words 76
4.4.3 Acoustic Vectors to Words 77
4.5 MSE and Discriminant Distance 78
4.6 Summary 79

II HYBRID HMM/MLP SYSTEMS 81

5 SPEECH RECOGNITION USING ANNs 83
5.1 Introduction 83
5.2 Fallacious Reasons for Using ANNs 85
5.3 Valid Reasons for Using ANNs 87
5.4 Neural Nets and Time Sequences 89
5.4.1 Static Networks with Buffered Input 90
5.4.2 Recurrent Networks 93
5.4.3 Partial Feedback of Context Units 94
5.4.4 Approximating Recurrent Networks by MLPs . . . 97
5.4.5 Discussion 99
5.5 ANN Models of HMMs 100
5.5.1 The Viterbi Network 100
5.5.2 The Alpha-Net 101
5.5.3 Combining ANNs and Dynamic Time Warping . . 101
5.5.4 ANNs for Nonlinear Transformations 104
5.5.5 ANNs for Preprocessing 105
5.6 Discrimination with Contextual MLPs 105
5.7 Summary 113

6 STATISTICAL INFERENCE IN MLPs 115
6.1 Introduction 115
6.2 ANNs and Statistical Inference 116
6.2.1 Discrete Case 117
6.2.2 Continuous Case 120
6.3 Recurrent MLP with Output Feedback 121
6.4 Practical Implications 125
6.4.1 Local Minima 125
6.4.2 Network Outputs Sum to One 125
6.4.3 Prior Class Probabilities and Likelihoods 126
6.4.4 Priors and MLP Output Biases 127
6.4.5 Conclusion 127
6.5 MLPs with Contextual Inputs 128
6.6 Classification of Acoustic Vectors 130
6.6.1 Experimental Approach 130
6.6.2 MLP Approach, Training and Cross-validation . . 131
6.6.3 MLP Results 133
6.6.4 Assessing Bayesian Properties of MLPs 133

6.6.5 Effect of Cross-Validation 136
6.6.6 Output Sigmoid Function 138
6.6.7 Feature Dependence 139
6.7 Radial Basis Functions 140
6.7.1 General Approach 140
6.7.2 RBFs and Tied Mixtures 142
6.7.3 RBFs for MAP Estimation 143
6.7.4 Lagrange Multipliers 145
6.7.5 Discussion 147
6.8 MLPs for Autoregressive Modeling 147
6.8.1 Linear Autoregressive Modeling 148
6.8.2 Predictive Neural Networks 149
6.8.3 Statistical Interpretation 149
6.8.4 Another Approach 150
6.8.5 Discussion 151
6.9 Summary 152

7 THE HYBRID HMM/MLP APPROACH 155
7.1 Introduction 155
7.2 Discriminant Markov Models 158
7.2.1 Formulation 158
7.2.2 Conditional Transition Probabilities 159
7.2.3 Maximum Likelihood Criterion 162
7.2.4 Viterbi Criterion 162
7.2.5 MLPs for Discriminant HMMs 163
7.3 Problem 165
7.4 Methods for Recognition at Word Level 166
7.4.1 MLP Training Methods 166
7.4.2 Posterior Probabilities and Likelihoods 168
7.4.3 Word Transition Costs 168
7.4.4 Segmentation of Training Data 170
7.4.5 Input Features 170
7.4.6 Better Speech Units and Phonological Rules 171
7.5 Word Recognition Results 171
7.6 Segmentation of training data 175
7.7 Resource Management (RM) task 177
7.7.1 Methods 177
7.7.2 Results 179
7.7.3 Discussion and Extensions 179

7.8 Discriminative Training and Priors 181
7.9 Summary 182

8 EXPERIMENTAL SYSTEMS 185
8.1 Introduction 185
8.2 Experiments on RM and TIMIT 187
8.2.1 Methods 187
8.2.2 Recognition Results 190
8.2.3 Discussion 193
8.3 Integrating the MLP into DECIPHER 195
8.3.1 Coming Full Circle: RM Experiments 196
8.3.2 Context-independent Models 197
8.4 Summary 199

9 CONTEXT-DEPENDENT MLPs 201
9.1 Introduction 201
9.2 CDNN: A Context-Dependent Neural Network 202
9.3 Theoretical Issue 203
9.4 Implementation Issue 207
9.5 Discussion and Results 209
9.5.1 The Unrestricted Split Net 209
9.5.2 The Topologically Restricted Net 210
9.5.3 Preliminary Results and Conclusion 210
9.6 Related Prior Work 211
9.7 Summary 212

10 SYSTEM TRADEOFFS 215
10.1 Introduction 215
10.2 Discrete HMM 216
10.3 Continuous-density HMM 218
10.4 Summary 221

11 TRAINING HARDWARE AND SOFTWARE 223
11.1 Introduction 223
11.2 Motivations 224
11.3 A Basic Neurocomputer Design - the RAP 226
11.3.1 The ICSI Ring Array Processor 226
11.3.2 Current Developments and Conclusions 228
11.4 Summary 229

III ADDITIONAL TOPICS 231

12 CROSS-VALIDATION IN MLP TRAINING 233
12.1 Introduction 233
12.2 Random Vector Problem 235
12.2.1 Methods 235
12.2.2 Results 236
12.3 Speech Recognition 239
12.3.1 Methods 239
12.3.2 Results 240
12.4 Summary 241

13 HMM/MLP AND PREDICTIVE MODELS 243
13.1 Introduction 243
13.2 Autoregressive HMMs 245
13.3 Full and Conditional Likelihoods 247
13.4 Gaussian Additive Noise 248
13.4.1 Training 248
13.4.2 Recognition 249
13.4.3 Discussion 250
13.5 Linear or Nonlinear AR Models? 250
13.6 ARCH Models 251
13.7 Summary 252

14 FEATURE EXTRACTION BY MLP 253
14.1 Introduction 253
14.2 MLP and Auto-Association 255
14.3 Explicit and Optimal Solution 257
14.4 Linear Hidden Units 259
14.5 Nonlinear Hidden Units 260
14.6 Experiments 261
14.7 Summary 263

IV FINALE 265

15 FINAL SYSTEM OVERVIEW 267
15.1 Introduction 267
15.2 System Description 267

15.2.1 Network Specifications 267
15.2.2 Training 269
15.2.3 Recognition 272
15.3 New Perspectives 273
15.4 Summary 274

16 CONCLUSIONS **275**
16.1 Introduction 275
16.2 Hybrid HMM/ANN Systems: Status 276
16.3 Future Research Issues 278
16.3.1 In Hybrid HMM/ANN Approaches for CSR 278
16.3.2 In General 279
16.4 Concluding Remark 280

Bibliography **281**

Index **307**

Acronyms **311**