

- 1 Minimum Classification Error (MCE) Approach in Pattern Recognition Wu Chou p. 1
- 1.1 Introduction p. 1
- 1.2 Optimal Classifier from Bayes Decision Theory p. 3
- 1.3 Discriminant Function Approach to Classifier Design p. 6
- 1.4 Speech Recognition and Hidden Markov Modeling p. 7
- 1.4.1 Hidden Markov Modeling of Speech p. 8
- 1.5 MCE Classifier Design Using Discriminant Functions p. 10
- 1.5.1 MCE Classifier Design Strategy p. 10
- 1.5.2 Optimization Methods p. 12
- 1.5.3 Other Optimization Methods p. 14
- 1.5.4 HMM as a Discriminant Function p. 15
- 1.5.5 Relation between MCE and MMI p. 18
- 1.5.6 Discussions and Comments p. 22
- 1.6 Embedded String Model Based MCE Training p. 24
- 1.6.1 String Model Based MCE Approach p. 25
- 1.6.2 Combined String Model Based MCE Approach p. 29
- 1.6.3 Discriminative Feature Extraction p. 32
- 1.7 Verification and Identification p. 33
- 1.7.1 Speaker Verification and Identification p. 35
- 1.7.2 Utterance Verification p. 37
- 1.8 Summary p. 40
- 2 Minimum Bayes-Risk Methods in Automatic Speech Recognition Vaibhava Goel and William Byrne p. 51
- 2.1 Minimum Bayes-Risk Classification Framework p. 52
- 2.1.1 Likelihood Ratio Based Hypothesis Testing p. 53
- 2.1.2 Maximum A-Posteriori Probability Classification p. 54
- 2.1.3 Previous Studies of Application Sensitive ASR p. 54
- 2.2 Practical MBR Procedures for ASR p. 55
- 2.2.1 Summation over Hidden State Sequences p. 56
- 2.2.2 MBR Recognition with N-best Lists p. 57
- 2.2.3 MBR Recognition with Lattices p. 57
- 2.3 Segmental MBR Procedures p. 64
- 2.3.1 Segmental Voting p. 66
- 2.3.2 ROVER p. 67
- 2.3.3 e-ROVER p. 68
- 2.4 Experimental Results p. 70
- 2.4.1 Parameter Tuning within the MBR Classification Rule p. 70
- 2.4.2 Utterance Level MBR Word and Keyword Recognition p. 72
- 2.4.3 ROVER and e-ROVER for Multilingual ASR p. 74
- 2.5 Summary p. 76
- 2.6 Acknowledgements p. 77
- 3 A Decision Theoretic Formulation for Robust Automatic Speech Recognition Qiang Huo p. 81
- 3.1 Introduction p. 81
- 3.2 Optimal Bayes' Decision Rule for ASR p. 83
- 3.3 Adaptive Decision Rules Constructed from Training Samples p. 85

- 3.3.1 Plug-in Bayes' Decision Rules with Maximum-likelihood Density Estimate p. 86
- 3.3.2 Maximum-Discriminant Decision Rules Minimizing the Empirical Classification Error p. 89
- 3.3.3 Discussion p. 90
- 3.4 Violations of Modeling Assumptions in ASR p. 91
- 3.4.1 Types of Distortions p. 91
- 3.4.2 Towards Adaptive and Robust ASR p. 92
- 3.5 Improving Adaptive Decision Rules via Decision Parameter Adaptation p. 93
- 3.5.1 Decision Parameter Adaptation for Stationary Operating Conditions p. 93
- 3.5.2 Decision Parameter Adaptation for Slowly Changing Operating Conditions p. 95
- 3.5.3 Decision Parameter Adaptation for Switching Operating Conditions p. 96
- 3.5.4 Discussion p. 97
- 3.6 Robust Decision Rules p. 97
- 3.6.1 Decision Rule Robustness p. 97
- 3.6.2 Minimax Classification Rule p. 98
- 3.6.3 Bayesian Predictive Classification Rule p. 100
- 3.6.4 Discussion p. 103
- 3.7 Summary p. 104
- 4 Speech Pattern Recognition using Neural Networks Shigeru Katagiri p. 115
- 4.1 Introduction p. 115
- 4.2 Bayes Decision Theory p. 117
- 4.2.1 Preparations p. 117
- 4.2.2 Decision Rule p. 117
- 4.2.3 Minimum Error-rate Classification p. 118
- 4.2.4 Probability Function Estimation p. 118
- 4.2.5 Discriminative Training p. 119
- 4.3 Speech Recognizers Based on Neural Networks p. 124
- 4.3.1 Preparations p. 124
- 4.3.2 Classification Error Minimization p. 125
- 4.3.3 Squared Error Minimization p. 130
- 4.3.4 Cross Entropy Minimization p. 133
- 4.4 Fusion of Multiple Classification Decisions p. 135
- 4.4.1 Principles p. 135
- 4.4.2 Examples of Embodiment p. 138
- 4.5 Concluding Remarks p. 143
- 4.6 Appendix: Maximizing Mutual Information p. 146
- 5 Large Vocabulary Speech Recognition Based on Statistical Methods Jean-Luc Gauvain and Lori Lamel p. 149
- 5.1 Introduction p. 149
- 5.2 Overview p. 150
- 5.3 Language Modeling p. 151
- 5.3.1 Text Preparation p. 153
- 5.3.2 Vocabulary Selection p. 154
- 5.3.3 N-gram Estimation p. 154
- 5.3.4 LM Adaptation p. 156
- 5.4 Pronunciation Modeling p. 156

- 5.5 Acoustic Modeling p. 159
- 5.5.1 Acoustic Front-end p. 159
- 5.5.2 Modeling Allophones p. 161
- 5.5.3 HMM Parameter Estimation p. 163
- 5.5.4 HMM Adaptation p. 165
- 5.6 Decoding p. 167
- 5.6.1 Speech/Non-speech Detection p. 168
- 5.6.2 Decoding Strategies p. 169
- 5.6.3 Efficiency p. 170
- 5.6.4 Confidence Measures p. 172
- 5.7 Indicative Performance Levels p. 172
- 5.7.1 Dictation p. 173
- 5.7.2 Speech Recognition for Dialog Systems p. 174
- 5.7.3 Transcription for Audio Indexation p. 175
- 5.8 Portability and Language Dependencies p. 177
- 6 Toward Spontaneous Speech Recognition and Understanding Sadaoki Furui p. 191
- 6.1 Introduction p. 191
- 6.2 Four Categories of Speech Recognition Tasks p. 193
- 6.3 Spontaneous Speech Recognition and Understanding - Review p. 195
- 6.3.1 Category I (human-to-human dialogue) p. 195
- 6.3.2 Category II (human-to-human monologue) p. 196
- 6.3.3 Category III (human-to-machine dialogue) p. 198
- 6.4 Japanese National Project on Spontaneous Speech Corpus and Processing Technology p. 200
- 6.4.1 Project Overview p. 200
- 6.4.2 Corpus p. 201
- 6.5 Automatic Transcription of Spontaneous Presentation p. 202
- 6.5.1 Recognition Task p. 202
- 6.5.2 Language and Acoustic Modeling p. 202
- 6.5.3 Recognition Results p. 203
- 6.5.4 Analysis on Individual Differences p. 205
- 6.5.5 Discussion p. 210
- 6.6 Automatic Speech Summarization and Evaluation p. 212
- 6.6.1 Summarization of Each Sentence Utterance p. 212
- 6.6.2 Summarization of Multiple Utterances p. 215
- 6.6.3 Evaluation p. 215
- 6.6.4 Discussion p. 218
- 6.7 Spontaneous Speech Recognition and Understanding Research Issues p. 219
- 6.7.1 Language Models and Corpora p. 219
- 6.7.2 Message-driven Speech Recognition and Understanding p. 220
- 6.7.3 Statistical Approaches and Speech Science p. 222
- 6.7.4 Research on the Human Brain p. 222
- 6.7.5 Dynamic Spectral Features p. 223
- 6.8 Conclusion p. 224
- 7 Speaker Authentication Qi Li and Biing-Hwang Juang p. 229
- 7.1 Introduction p. 229

- 7.1.1 Speaker Recognition and Verification p. 230
- 7.1.2 Verbal Information Verification p. 232
- 7.2 Pattern Recognition in Speaker Authentication p. 234
- 7.2.1 Bayesian Decision Theory p. 234
- 7.2.2 Stochastic Models for Stationary Process p. 236
- 7.2.3 Stochastic Models for Non-Stationary Process p. 237
- 7.2.4 Speech Segmentation p. 239
- 7.2.5 Statistical Verification p. 239
- 7.3 Speaker Verification System p. 240
- 7.4 Verbal Information Verification p. 243
- 7.4.1 Utterance Segmentation p. 244
- 7.4.2 Subword Hypothesis Testing p. 244
- 7.4.3 Confidence Measure Calculation p. 245
- 7.4.4 Sequential Utterance Verification p. 247
- 7.4.5 VIV Experimental Results p. 249
- 7.5 Speaker Authentication by Combining SV and VIV p. 251
- 7.6 Summary p. 254
- 8 HMMs for Language Processing Problems Richard M. Schwartz and John Makhoul p. 261
- 8.1 Introduction p. 261
- 8.2 Use of Probabilities p. 262
- 8.2.1 Hidden Markov Models p. 263
- 8.3 Name Spotting p. 264
- 8.4 Topic Classification p. 266
- 8.4.1 The Model p. 267
- 8.4.2 Estimating HMM Parameters p. 268
- 8.4.3 Classification p. 269
- 8.4.4 Experiments p. 269
- 8.5 Information Retrieval p. 270
- 8.5.1 A Bayesian Model for IR p. 270
- 8.5.2 Training the IR HMM p. 271
- 8.5.3 Performance p. 271
- 8.6 Event Tracking p. 272
- 8.7 Unsupervised Topic Detection p. 273
- 8.8 Summary p. 275
- 9 Statistical Language Models With Embedded Latent Semantic Knowledge Jerome R. Bellegarda p. 277
- 9.1 Introduction p. 277
- 9.1.1 Scope Locality p. 277
- 9.1.2 Syntactically-Driven Span Extension p. 278
- 9.1.3 Semantically-Driven Span Extension p. 279
- 9.1.4 Organization p. 279
- 9.2 Latent Semantic Analysis p. 280
- 9.2.1 Feature Extraction p. 280
- 9.2.2 Singular Value Decomposition p. 281
- 9.2.3 General Behavior p. 282

- 9.3 LSA Feature Space p. 283
- 9.3.1 Word Clustering p. 284
- 9.3.2 Word Cluster Example p. 284
- 9.3.3 Document Clustering p. 285
- 9.3.4 Document Cluster Example p. 286
- 9.4 Semantic Classification p. 287
- 9.4.1 Framework Extension p. 287
- 9.4.2 Semantic Inference p. 288
- 9.4.3 Caveats p. 289
- 9.5 N-gram+LSA Language Modeling p. 290
- 9.5.1 LSA Component p. 290
- 9.5.2 Integration with N-grams p. 292
- 9.5.3 Context Scope Selection p. 294
- 9.6 Smoothing p. 295
- 9.6.1 Word Smoothing p. 295
- 9.6.2 Document Smoothing p. 296
- 9.6.3 Joint Smoothing p. 297
- 9.7 Experiments p. 297
- 9.7.1 Experimental Conditions p. 298
- 9.7.2 Experimental Results p. 298
- 9.7.3 Context Scope Selection p. 299
- 9.8 Inherent Trade-Offs p. 300
- 9.8.1 Cross-Domain Training p. 300
- 9.8.2 Discussion p. 301
- 9.9 Conclusion p. 302
- 10 Semantic Information Processing of Spoken Language - How May I Help You?sm A. L. Gorin and A. Abella and T. Alonso and G. Riccardi and J. H. Wright p. 309
- 10.1 Introduction p. 309
- 10.2 Call-Classification p. 311
- 10.3 Language Modeling for Recognition and Understanding p. 316
- 10.4 Dialog p. 318
- 10.5 Conclusions p. 319
- 11 Machine Translation Using Statistical Modeling Herman Ney and F. J. Och p. 321
- 11.1 Introduction p. 321
- 11.2 Statistical Decision Theory and Linguistics p. 323
- 11.2.1 The Statistical Approach p. 323
- 11.2.2 Bayes Decision Rule for Written Language Translation p. 324
- 11.2.3 Related Approaches p. 324
- 11.3 Alignment and Lexicon Models p. 325
- 11.3.1 Concept of Alignment Modelling p. 325
- 11.3.2 Hidden Markov Models p. 326
- 11.3.3 Models IBM 1-5 p. 329
- 11.3.4 Training p. 331
- 11.3.5 Search p. 332

- 11.3.6 Algorithmic Differences between Speech Recognition and Language Translation p. 334
- 11.4 Alignment Templates: From Single Words to Word Groups p. 335
- 11.4.1 Concept p. 335
- 11.4.2 Training p. 337
- 11.4.3 Search p. 338
- 11.5 Experimental Results p. 338
- 11.5.1 The Task and the Corpus p. 338
- 11.5.2 Offline Results p. 339
- 11.5.3 Integration into the Verbmobil Prototype System p. 341
- 11.5.4 Final Evaluation p. 341
- 11.6 Speech Translation: The Integrated Approach p. 343
- 11.6.1 Principle p. 343
- 11.6.2 Practical Implementation p. 344
- 11.7 Summary p. 346
- 11.8 References p. 347
- 12 Modeling Topics for Detection and Tracking James Allan p. 353
- 12.1 Topic Detection and Tracking p. 353
- 12.1.1 Topic and Events p. 354
- 12.1.2 TDT Tasks p. 354
- 12.1.3 Corpora p. 356
- 12.1.4 Evaluation p. 357
- 12.2 Basic Topic Models p. 359
- 12.2.1 Vector Space p. 359
- 12.2.2 Language Models p. 360
- 12.3 Implementing the Models p. 360
- 12.3.1 Named Entities p. 361
- 12.3.2 Document Expansion p. 361
- 12.3.3 Clustering p. 362
- 12.3.4 Time Decay p. 363
- 12.4 Comparing Models p. 363
- 12.4.1 Nearest Neighbors p. 363
- 12.4.2 Decision Trees p. 364
- 12.4.3 Model-to-Model p. 365
- 12.5 Miscellaneous Issues p. 365
- 12.5.1 Deferral p. 366
- 12.5.2 Multi-modal Issues p. 366
- 12.5.3 Multi-lingual Issues p. 367
- 12.6 Using TDT Interactively p. 368
- 12.6.1 Demonstrations p. 368
- 12.6.2 Timelines p. 369
- 12.7 Modeling Events p. 370
- 12.8 Conclusion p. 371
- Index p. 377