
Contents

1	Introduction	1
1.1	Thematic Context	3
1.2	Functional Principles of Markov Models	4
1.3	Goal and Structure of the Book	5
2	Application Areas	9
2.1	Speech	9
2.2	Writing	15
2.3	Biological Sequences	24
2.4	Outlook	28
Part I Theory		
3	Foundations of Mathematical Statistics	35
3.1	Random Experiment, Event, and Probability	35
3.2	Random Variables and Probability Distributions	37
3.3	Parameters of Probability Distributions	39
3.4	Normal Distributions and Mixture Models	41
3.5	Stochastic Processes and Markov Chains	42
3.6	Principles of Parameter Estimation	44
3.6.1	Maximum Likelihood Estimation	44
3.6.2	Maximum a Posteriori Estimation	47
3.7	Bibliographical Remarks	49
4	Vector Quantization and Mixture Estimation	51
4.1	Definition	52
4.2	Optimality	53
4.2.1	Nearest-Neighbor Condition	54
4.2.2	Centroid Condition	55
4.3	Algorithms for Vector Quantizer Design	57
4.3.1	Lloyd's Algorithm	57
4.3.2	LBG Algorithm	59

4.3.3	<i>k</i> -Means Algorithm	61
4.4	Estimation of Mixture Density Models	62
4.4.1	EM Algorithm	63
4.4.2	EM Algorithm for Gaussian Mixtures	66
4.5	Bibliographical Remarks	69
5	Hidden Markov Models	71
5.1	Definition	72
5.2	Modeling Outputs	73
5.3	Use Cases	75
5.4	Notation	78
5.5	Evaluation	78
5.5.1	The Total Output Probability	79
5.5.2	Forward Algorithm	80
5.5.3	The Optimal Output Probability	82
5.6	Decoding	85
5.6.1	Viterbi Algorithm	86
5.7	Parameter Estimation	87
5.7.1	Foundations	88
5.7.2	Forward–Backward Algorithm	89
5.7.3	Training Methods	90
5.7.4	Baum–Welch Algorithm	92
5.7.5	Viterbi Training	96
5.7.6	Segmental <i>k</i> -Means Algorithm	100
5.7.7	Multiple Observation Sequences	103
5.8	Model Variants	104
5.8.1	Alternative Algorithms	104
5.8.2	Alternative Model Architectures	104
5.9	Bibliographical Remarks	105
6	<i>n</i>-Gram Models	107
6.1	Definition	107
6.2	Use Cases	109
6.3	Notation	110
6.4	Evaluation	110
6.5	Parameter Estimation	114
6.5.1	Redistribution of Probability Mass	115
6.5.2	Discounting	115
6.5.3	Incorporation of More General Distributions	117
6.5.4	Interpolation	118
6.5.5	Backing off	120
6.5.6	Optimization of Generalized Distributions	121
6.6	Model Variants	123
6.6.1	Category-Based Models	123
6.6.2	Longer Temporal Dependencies	126
6.7	Bibliographical Remarks	126

Part II Practice

7	Computations with Probabilities	133
7.1	Logarithmic Probability Representation	134
7.2	Lower Bounds for Probabilities	137
7.3	Codebook Evaluation for Semi-continuous HMMs	138
7.4	Probability Ratios	139
8	Configuration of Hidden Markov Models	143
8.1	Model Topologies	143
8.2	Modularization	144
8.2.1	Context-Independent Sub-word Units	145
8.2.2	Context-Dependent Sub-word Units	146
8.3	Compound Models	147
8.4	Profile HMMs	149
8.5	Modeling Outputs	151
9	Robust Parameter Estimation	153
9.1	Feature Optimization	155
9.1.1	Decorrelation	156
9.1.2	Principal Component Analysis I	157
9.1.3	Whitening	162
9.1.4	Dimensionality Reduction	163
9.1.5	Principal Component Analysis II	164
9.1.6	Linear Discriminant Analysis	165
9.2	Tying	169
9.2.1	Sub-model Units	170
9.2.2	State Tying	174
9.2.3	Tying in Mixture Models	178
9.3	Initialization of Parameters	181
10	Efficient Model Evaluation	183
10.1	Efficient Evaluation of Mixture Densities	184
10.2	Efficient Decoding of Hidden Markov Models	185
10.2.1	Beam Search Algorithm	186
10.3	Efficient Generation of Recognition Results	189
10.3.1	First-Best Decoding of Segmentation Units	189
10.3.2	Algorithms for N -Best Search	190
10.4	Efficient Parameter Estimation	192
10.4.1	Forward–Backward Pruning	192
10.4.2	Segmental Baum–Welch Algorithm	193
10.4.3	Training of Model Hierarchies	194
10.5	Tree-Like Model Organization	195
10.5.1	HMM Prefix Trees	195
10.5.2	Tree-Like Representation for n -Gram Models	197

11	Model Adaptation	201
11.1	Basic Principles	201
11.2	Adaptation of Hidden Markov Models	202
11.2.1	Maximum-Likelihood Linear-Regression	204
11.3	Adaptation of n -Gram Models	207
11.3.1	Cache Models	207
11.3.2	Dialog-Step Dependent Models	208
11.3.3	Topic-Based Language Models	208
12	Integrated Search Methods	211
12.1	HMM Networks	215
12.2	Multi-pass Search	216
12.3	Search Space Copies	217
12.3.1	Context-Based Search Space Copies	218
12.3.2	Time-Based Search Space Copies	219
12.3.3	Language-Model Look-Ahead	220
12.4	Time-Synchronous Parallel Model Decoding	222
12.4.1	Generation of Segment Hypotheses	222
12.4.2	Language-Model-Based Search	223
Part III Systems		
13	Speech Recognition	229
13.1	Recognition System of RWTH Aachen University	229
13.1.1	Feature Extraction	230
13.1.2	Acoustic Modeling	230
13.1.3	Language Modeling	231
13.1.4	Search	231
13.2	BBN Speech Recognizer BYBLOS	231
13.2.1	Feature Extraction	232
13.2.2	Acoustic Modeling	232
13.2.3	Language Modeling	232
13.2.4	Search	233
13.3	ESMERALDA	233
13.3.1	Feature Extraction	234
13.3.2	Acoustic Modeling	235
13.3.3	Statistical and Declarative Language Modeling	235
13.3.4	Incremental Search	236
14	Handwriting Recognition	237
14.1	Recognition System by BBN	238
14.1.1	Preprocessing	238
14.1.2	Feature Extraction	238
14.1.3	Script Modeling	239
14.1.4	Language Modeling and Search	240
14.2	Recognition System of RWTH Aachen University	240
14.2.1	Preprocessing	240

14.2.2	Feature Extraction	241
14.2.3	Script Modeling	241
14.2.4	Language Modeling and Search	242
14.3	ESMERALDA Offline Recognition System	242
14.3.1	Preprocessing	242
14.3.2	Feature Extraction	243
14.3.3	Handwriting Model	244
14.3.4	Language Modeling and Search	244
14.4	Bag-of-Features Hidden Markov Models	244
15	Analysis of Biological Sequences	249
15.1	HMMER	250
15.1.1	Model Structure	250
15.1.2	Parameter Estimation	250
15.1.3	Interoperability	251
15.2	SAM	251
15.3	ESMERALDA	252
15.3.1	Feature Extraction	252
15.3.2	Statistical Models of Proteins	253
	References	255
	Index	273

Markov Models for Pattern Recognition

From Theory to Applications

Fink, G.A.

2014, XIII, 276 p. 45 illus., Hardcover

ISBN: 978-1-4471-6307-7