

Finn V. Jensen and Thomas D. Nielsen

Bayesian Networks and Decision Graphs

February 8, 2007

Springer

Berlin Heidelberg New York

Hong Kong London

Milan Paris Tokyo

Table of Contents

Preface	v
1 Prerequisites on Probability Theory	1
1.1 Two Perspectives on Probability Theory	1
1.2 Fundamentals of Probability Theory	2
1.2.1 Conditional Probabilities	4
1.2.2 Probability Calculus	5
1.2.3 Conditional Independence	6
1.3 Probability Calculus for Variables	7
1.3.1 Calculations with Probability Tables: An Example	11
1.4 An Algebra of Potentials	13
1.5 Random Variables	15
1.5.1 Continuous Distributions	15
1.6 Exercises	16

Part I Probabilistic Graphical Models

2 Causal and Bayesian Networks	23
2.1 Reasoning Under Uncertainty	23
2.1.1 Car Start Problem	23
2.1.2 A Causal Perspective on the Car Start Problem	24
2.2 Causal Networks and d-Separation	26
2.2.1 d-separation	30
2.3 Bayesian Networks	32
2.3.1 Definition of Bayesian Networks	32
2.3.2 The Chain Rule for Bayesian Networks	35
2.3.3 Inserting Evidence	39
2.3.4 Calculating Probabilities in Practice	41
2.4 Graphical Models – Formal Languages for Model Specification	42
2.5 Summary	44

2.6	Bibliographical Notes	45
2.7	Exercises	45
3	Building Models	51
3.1	Catching the Structure	51
3.1.1	Milk Test	52
3.1.2	Cold or Angina?	54
3.1.3	Insemination	55
3.1.4	A Simplified Poker Game	57
3.1.5	Naive Bayes Models	58
3.1.6	Causality	60
3.2	Determining the Conditional Probabilities	60
3.2.1	Milk Test	60
3.2.2	Stud Farm	62
3.2.3	Poker Game	66
3.2.4	Transmission of Symbol Strings	68
3.2.5	Cold or Angina?	71
3.2.6	Why Causal Networks?	72
3.3	Modeling Methods	73
3.3.1	Undirected Relations	73
3.3.2	Noisy-Or	75
3.3.3	Divorcing	78
3.3.4	Noisy Functional Dependence	80
3.3.5	Expert Disagreements	81
3.3.6	Object-Oriented Bayesian Networks	84
3.3.7	Dynamic Bayesian Networks	91
3.3.8	How to Deal with Continuous Variables	93
3.3.9	Interventions	96
3.4	Special Features	97
3.4.1	Joint Probability Tables	98
3.4.2	Most-Probable Explanation	98
3.4.3	Data Conflict	98
3.4.4	Sensitivity Analysis	99
3.5	Summary	100
3.6	Bibliographical Notes	101
3.7	Exercises	102
4	Belief Updating in Bayesian Networks	109
4.1	Introductory Examples	109
4.1.1	A Single Marginal	110
4.1.2	Different Evidence Scenarios	111
4.1.3	All Marginals	114
4.2	Graph-Theoretic Representation	115
4.2.1	Task and Notation	115
4.2.2	Domain Graphs	116

4.3	Triangulated Graphs and Join Trees	119
4.3.1	Join Trees	122
4.4	Propagation in Junction Trees	124
4.4.1	Lazy Propagation in Junction Trees	127
4.5	Exploiting the Information Scenario	130
4.5.1	Barren Nodes	130
4.5.2	d-Separation	131
4.6	Nontriangulated Domain Graphs	132
4.6.1	Triangulation of Graphs	134
4.6.2	Triangulation of Dynamic Bayesian Networks	137
4.7	Exact Propagation with Bounded Space	140
4.7.1	Recursive Conditioning	140
4.8	Stochastic Simulation in Bayesian Networks	145
4.8.1	Probabilistic Logic Sampling	146
4.8.2	Likelihood Weighting	148
4.8.3	Gibbs Sampling	150
4.9	Loopy Belief Propagation	152
4.10	Summary	154
4.11	Bibliographical Notes	156
4.12	Exercises	157
5	Analysis Tools for Bayesian Networks	167
5.1	IEJ Trees	168
5.2	Joint Probabilities and A-Saturated Junction Trees	169
5.2.1	A-Saturated Junction Trees	169
5.3	Configuration of Maximal Probability	171
5.4	Axioms for Propagation in Junction Trees	173
5.5	Data Conflict	174
5.5.1	Insemination	175
5.5.2	The Conflict Measure conf	175
5.5.3	Conflict or Rare Case	176
5.5.4	Tracing of Conflicts	177
5.5.5	Other Approaches to Conflict Detection	179
5.6	SE Analysis	179
5.6.1	Example and Definitions	179
5.6.2	h -Saturated Junction Trees and SE Analysis	182
5.7	Sensitivity to Parameters	184
5.7.1	One-Way Sensitivity Analysis	187
5.7.2	Two-Way Sensitivity Analysis	188
5.8	Summary	188
5.9	Bibliographical Notes	190
5.10	Exercises	191

6	Parameter Estimation	195
6.1	Complete Data	195
6.1.1	Maximum Likelihood Estimation	196
6.1.2	Bayesian Estimation	197
6.2	Incomplete Data	200
6.2.1	Approximate Parameter Estimation: The EM Algorithm	201
6.2.2	*Why We Cannot Perform Exact Parameter Estimation	207
6.3	Adaptation	207
6.3.1	Fractional Updating	210
6.3.2	Fading	211
6.3.3	*Specification of an Initial Sample Size	212
6.3.4	Example: Strings of Symbols	213
6.3.5	Adaptation to Structure	214
6.3.6	*Fractional Updating as an Approximation	215
6.4	Tuning	218
6.4.1	Example	220
6.4.2	Determining $\text{grad dist}(x, y)$ as a Function of t	222
6.5	Summary	223
6.6	Bibliographical Notes	225
6.7	Exercises	226
7	Learning the Structure of Bayesian Networks	229
7.1	Constraint-Based Learning Methods	230
7.1.1	From Skeleton to DAG	231
7.1.2	From Independence Tests to Skeleton	234
7.1.3	Example	235
7.1.4	Constraint-Based Learning on Data Sets	237
7.2	Ockham's Razor	240
7.3	Score-Based Learning	241
7.3.1	Score Functions	242
7.3.2	Search Procedures	245
7.3.3	Chow-Liu Trees	250
7.3.4	*Bayesian Score Functions	253
7.4	Summary	258
7.5	Bibliographical Notes	260
7.6	Exercises	261
8	Bayesian Networks as Classifiers	265
8.1	Naive Bayes Classifiers	266
8.2	Evaluation of Classifiers	268
8.3	Extensions of Naive Bayes Classifiers	270
8.4	Classification Trees	272
8.5	Summary	274
8.6	Bibliographical Notes	275
8.7	Exercises	276

Part II Decision Graphs

9	Graphical Languages for Specification of Decision Problems	279
9.1	One-Shot Decision Problems	280
9.1.1	Fold or Call?	281
9.1.2	Mildew	282
9.1.3	One Decision in General	283
9.2	Utilities	284
9.2.1	Instrumental Rationality	287
9.3	Decision Trees	290
9.3.1	A Couple of Examples	293
9.3.2	Coalesced Decision Trees	295
9.3.3	Solving Decision Trees	296
9.4	Influence Diagrams	302
9.4.1	Extended Poker Model	302
9.4.2	Definition of Influence Diagrams	305
9.4.3	Repetitive Decision Problems	308
9.5	Asymmetric Decision Problems	310
9.5.1	Different Sources of Asymmetry	314
9.5.2	Unconstrained Influence Diagrams	316
9.5.3	Sequential Influence Diagrams	322
9.6	Decision Problems with Unbounded Time Horizons	324
9.6.1	Markov Decision Processes	324
9.6.2	Partially Observable Markov Decision Processes	330
9.7	Summary	332
9.8	Bibliographical Notes	337
9.9	Exercises	337
10	Solution Methods for Decision Graphs	343
10.1	Solutions to Influence Diagrams	343
10.1.1	The Chain Rule for Influence Diagrams	345
10.1.2	Strategies and Expected Utilities	346
10.1.3	An Example	352
10.2	Variable Elimination	353
10.2.1	Strong Junction Trees	355
10.2.2	Required Past	358
10.2.3	Policy Networks	360
10.3	Node Removal and Arc Reversal	362
10.3.1	Node Removal	362
10.3.2	Arc Reversal	363
10.3.3	An Example	365
10.4	Solutions to Unconstrained Influence Diagrams	367
10.4.1	Minimizing the S-DAG	367
10.4.2	Determining Policies and Step Functions	371

10.5 Decision Problems Without a Temporal Ordering:	
Troubleshooting	373
10.5.1 Action Sequences	373
10.5.2 A Greedy Approach	375
10.5.3 Call Service	378
10.5.4 Questions	378
10.6 Solutions to Decision Problems with Unbounded Time Horizon	380
10.6.1 A Basic Solution	380
10.6.2 Value Iteration	381
10.6.3 Policy Iteration	385
10.6.4 Solving Partially Observable Markov Decision Processes*	388
10.7 Limited Memory Influence Diagrams	392
10.8 Summary	395
10.9 Bibliographical Notes	400
10.10 Exercises	401
11 Methods for Analyzing Decision Problems	407
11.1 Value of Information	407
11.1.1 Test for Infected Milk?	407
11.1.2 Myopic Hypothesis-Driven Data Request	409
11.1.3 Non-Utility-Based Value Functions	411
11.2 Finding the Relevant Past and Future of a Decision Problem . .	413
11.2.1 Identifying the Required Past	415
11.3 Sensitivity Analysis	420
11.3.1 Example	421
11.3.2 One-Way Sensitivity Analysis in General	423
11.4 Summary	426
11.5 Bibliographical Notes	427
11.6 Exercises	427
List of Notation	429
References	431
Index	441