

Wendt

Q.4

76,76

E95

P755

1999

Robert G. Cowell
A. Philip Dawid
Steffen L. Lauritzen
David J. Spiegelhalter

Probabilistic Networks and Expert Systems

With 45 Illustrations



Springer

Contents

Preface	v
1 Introduction	1
1.1 What is this book about?	1
1.2 What is in this book?	2
1.3 What is not in this book?	3
1.4 How should this book be used?	4
2 Logic, Uncertainty, and Probability	5
2.1 What is an expert system?	5
2.2 Diagnostic decision trees	6
2.3 Production systems	7
2.4 Coping with uncertainty	8
2.5 The naïve probabilistic approach	10
2.6 Interpretations of probability	11
2.7 Axioms	13
2.8 Bayes' theorem	14
2.9 Bayesian reasoning in expert systems	17
2.10 A broader context for probabilistic expert systems	21
3 Building and Using Probabilistic Networks	25
3.1 Graphical modelling of the domain	26
3.1.1 Qualitative modelling	27
3.1.2 Probabilistic modelling	28

3.1.3	Quantitative modelling	29
3.1.4	Further background to the elicitation process	29
3.2	From specification to inference engine	31
3.2.1	Moralization	31
3.2.2	From moral graph to junction tree	33
3.3	The inference process	34
3.3.1	The clique-marginal representation	36
3.3.2	Incorporation of evidence	36
3.4	Bayesian networks as expert systems	37
3.5	Background references and further reading	40
3.5.1	Structuring the graph	40
3.5.2	Specifying the probability distribution	40
4	Graph Theory	43
4.1	Basic concepts	43
4.2	Chordal and decomposable graphs	49
4.3	Junction trees	52
4.4	From chain graph to junction tree	55
4.4.1	Triangulation	57
4.4.2	Elimination tree	59
4.5	Background references and further reading	61
5	Markov Properties on Graphs	63
5.1	Conditional independence	63
5.2	Markov fields over undirected graphs	66
5.3	Markov properties on directed acyclic graphs	70
5.4	Markov properties on chain graphs	75
5.5	Current research directions	79
5.5.1	Markov equivalence	79
5.5.2	Other graphical representations	80
5.6	Background references and further reading	80
6	Discrete Networks	83
6.1	An illustration of local computation	84
6.2	Definitions	85
6.2.1	Basic operations	86
6.3	Local computation on the junction tree	87
6.3.1	Graphical specification	87
6.3.2	Numerical specification and initialization	87
6.3.3	Charges	88
6.3.4	Flow of information between adjacent cliques	88
6.3.5	Active flows	89
6.3.6	Reaching equilibrium	90
6.3.7	Scheduling of flows	92
6.3.8	Two-phase propagation	92

6.3.9	Entering and propagating evidence	93
6.3.10	A propagation example	95
6.4	Generalized marginalization operations	95
6.4.1	Maximization	97
6.4.2	Degeneracy of the most probable configuration . . .	99
6.4.3	Simulation	99
6.4.4	Finding the M most probable configurations	101
6.4.5	Sampling without replacement	103
6.4.6	Fast retraction	104
6.4.7	Moments of functions	106
6.5	Example: CH-ASIA	109
6.5.1	Description	109
6.5.2	Graphical specification	109
6.5.3	Numerical specification	109
6.5.4	Initialization	112
6.5.5	Propagation without evidence	114
6.5.6	Propagation with evidence	114
6.5.7	Max-propagation	119
6.6	Dealing with large cliques	120
6.6.1	Truncating small numbers	121
6.6.2	Splitting cliques	122
6.7	Current research directions and further reading	123
7	Gaussian and Mixed Discrete-Gaussian Networks	125
7.1	CG distributions	126
7.2	Basic operations on CG potentials	127
7.3	Marked graphs and their junction trees	131
7.3.1	Decomposition of marked graphs	131
7.3.2	Junction trees with strong roots	133
7.4	Model specification	135
7.5	Operating in the junction tree	137
7.5.1	Initializing the junction tree	138
7.5.2	Charges	138
7.5.3	Entering evidence	139
7.5.4	Flow of information between adjacent cliques	139
7.5.5	Two-phase propagation	141
7.6	A simple Gaussian example	143
7.7	Example: WASTE	144
7.7.1	Structural specification	145
7.7.2	Numerical specification	146
7.7.3	Strong triangulation	147
7.7.4	Forming the junction tree	148
7.7.5	Initializing the junction tree	148
7.7.6	Entering evidence	149
7.8	Complexity considerations	150

7.9	Numerical instability problems	151
7.9.1	Exact marginal densities	152
7.10	Current research directions	152
7.11	Background references and further reading	153
8	Discrete Multistage Decision Networks	155
8.1	The nature of multistage decision problems	156
8.2	Solving the decision problem	157
8.3	Decision potentials	159
8.4	Network specification and solution	163
8.4.1	Structural and numerical specification	163
8.4.2	Causal consistency lemma	165
8.4.3	Making the elimination tree	166
8.4.4	Initializing the elimination tree	167
8.4.5	Message passing in the elimination tree	168
8.4.6	Proof of elimination tree solution	169
8.5	Example: OIL WILDCATTER	172
8.5.1	Specification	172
8.5.2	Making the elimination tree	175
8.5.3	Initializing the elimination tree	176
8.5.4	Collecting evidence	177
8.6	Example: DEC-ASIA	177
8.7	Triangulation issues	183
8.8	Asymmetric problems	184
8.9	Background references and further reading	187
9	Learning About Probabilities	189
9.1	Statistical modelling and parameter learning	189
9.2	Parametrizing a directed Markov model	190
9.3	Maximum likelihood with complete data	192
9.4	Bayesian updating with complete data	193
9.4.1	Priors for DAG models	193
9.4.2	Specifying priors: An example	197
9.4.3	Updating priors with complete data: An example	199
9.5	Incomplete data	200
9.5.1	Sequential and batch methods	201
9.6	Maximum likelihood with incomplete data	202
9.6.1	The EM algorithm	202
9.6.2	Penalized EM algorithm	204
9.7	Bayesian updating with incomplete data	204
9.7.1	Exact theory	206
9.7.2	Retaining global independence	207
9.7.3	Retaining local independence	209
9.7.4	Reducing the mixtures	211

9.7.5	Simulation results: full mixture reduction	213
9.7.6	Simulation results: partial mixture reduction	214
9.8	Using Gibbs sampling for learning	216
9.9	Hyper Markov laws for undirected models	221
9.10	Current research directions and further reading	222
10	Checking Models Against Data	225
10.1	Scoring rules	226
10.1.1	Standardization	227
10.2	Parent-child monitors	229
10.2.1	Batch monitors	232
10.2.2	Missing data	233
10.3	Node monitors	234
10.4	Global monitors	235
10.4.1	Example: CHILD	236
10.5	Simulation experiments	238
10.6	Further reading	241
11	Structural Learning	243
11.1	Purposes of modelling	244
11.2	Inference about models	244
11.3	Criteria for comparing models	245
11.3.1	Maximized likelihood	246
11.3.2	Predictive assessment	247
11.3.3	Marginal likelihood	248
11.3.4	Model probabilities	249
11.3.5	Model selection and model averaging	250
11.4	Graphical models and conditional independence	251
11.5	Classes of models	253
11.5.1	Models containing only observed quantities	253
11.5.2	Models with latent or hidden variables	254
11.5.3	Missing data	255
11.6	Handling multiple models	256
11.6.1	Search strategies	256
11.6.2	Probability specification	258
11.6.3	Prior information on parameters	260
11.6.4	Variable precision	261
	Epilogue	265
A	Conjugate Analysis for Discrete Data	267
A.1	Bernoulli process	267
A.2	Multinomial process	269
B	Gibbs Sampling	271

B.1	Gibbs sampling	271
B.2	Sampling from the moral graph	273
B.3	General probability densities	274
B.4	Further reading	275
C	Information and Software on the World Wide Web	277
C.1	Information about probabilistic networks	277
C.2	Software for probabilistic networks	279
C.3	Markov chain Monte Carlo methods	280
	Bibliography	281
	Author Index	307
	Subject Index	313

ILLUSTRATION SAMPLE

under the re-representation of P as factorizing over \mathcal{D}^m , some of the conditional independence properties displayed in the original DAG \mathcal{D} , such as the one between $n5$ and $n6$ given $n2$ in Figure 3.1, may lose their representation in graphical form. They still hold, but are now effectively buried in the quantitative component of the model. Only those conditional independence properties that retain a representation in the graph \mathcal{D}^m are exploited in the further analysis.

3.2.2 From moral graph to junction tree

The passage from moral graph to junction tree proceeds in the same way, regardless of the type of the initial graph. The first stage is to add sufficient edges to the moral graph, to make the resulting graph $(\mathcal{K}^m)'$ *triangulated* or *chordal*. Algorithms to do this are discussed in Section 4.4 below. The result is shown in Figure 3.4. The joint distribution, which factorizes on the

:

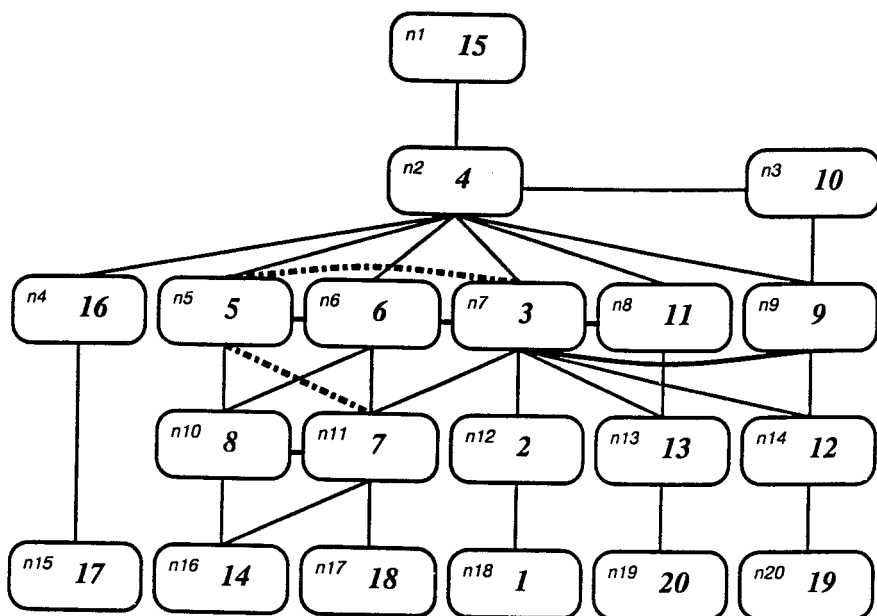


FIGURE 3.4. A triangulation of the moral graph of Figure 3.2, showing a perfect ordering of the nodes arising from maximum cardinality search (Algorithm 4.9).

moral graph \mathcal{K}^m , will also factorize on the larger triangulated graph $(\mathcal{K}^m)'$, since each clique in the moral graph will be either a clique or a subset of a clique in the triangulated graph. Hence, by a simple rearrangement of terms, we can transform the joint density into a product of factors on the