

Frontiers
in
Artificial
Intelligence
and
Applications

APPROXIMATION METHODS FOR EFFICIENT LEARNING OF BAYESIAN NETWORKS

Carsten Riggelsen

IOS
Press

Contents

Foreword	ix
1. INTRODUCTION	1
2. PRELIMINARIES	5
1 Random variables and conditional independence	5
2 Graph theory	6
3 Markov properties	7
3.1 The Markov blanket	11
3.2 Equivalence of DAGs	12
4 Bayesian networks	14
5 Bayesian network specification	14
5.1 Bayesian parameter specification	15
3. LEARNING BAYESIAN NETWORKS FROM DATA	19
1 The basics	19
2 Learning parameters	20
2.1 The Maximum-Likelihood approach	20
2.2 The Bayesian approach	21
3 Learning models	24
3.1 The penalised likelihood approach	25
3.2 The Bayesian approach	26
3.2.1 Learning via the marginal likelihood	26
3.2.2 Determining the hyper parameter	29
3.3 Marginal and penalised likelihood	32

3.4	Search methodologies	33
3.4.1	Model search space	34
3.4.2	Traversal strategy	35
4.	MONTE CARLO METHODS AND MCMC SIMULATION	37
1	Monte Carlo methods	37
1.1	Importance sampling	38
1.1.1	Choice of the sampling distribution	40
2	Markov chain Monte Carlo—MCMC	41
2.1	Markov chains	42
2.1.1	The invariant target distribution	43
2.1.2	Reaching the invariant distribution	43
2.1.3	Metropolis-Hastings sampling	44
2.1.4	Gibbs sampling	46
2.1.5	Mixing, burn-in and convergence of MCMC	49
2.1.6	The importance of blocking	51
3	Learning models via MCMC	53
3.1	Sampling models	54
3.2	Sampling edges	54
3.3	Blocking edges	56
3.3.1	Blocks and Markov blankets	58
3.3.2	Sampling blocks	59
3.3.3	Validity of the sampler	62
3.3.4	The MB-MCMC model sampler	62
3.3.5	Evaluation	64
3.3.6	Conclusion	69
5.	LEARNING FROM INCOMPLETE DATA	71
1	The concept of incomplete data	72
1.1	Missing data mechanisms	73
2	Learning from incomplete data	75
2.1	Likelihood decomposition	76
2.1.1	Complications for learning parameters	78
2.1.2	Bayesian sequential updating	79
2.1.3	Complications for learning models	80
3	Principled iterative methods	81
3.1	Expectation Maximisation—EM	81

3.1.1	Structural EM—SEM	85
3.2	Data Augmentation—DA	87
3.2.1	DA and eliminating the P-step—DA-P	90
3.3	DA-P and model learning—MDA-P	92
3.4	Efficiency issues of MDA-P	93
3.4.1	Properties of the sub-MCMC samplers	93
3.4.2	Interdependence between samplers	95
3.5	Imputation via importance sampling	97
3.5.1	The general idea	97
3.5.2	Importance sampling in the I-step—ISMDA-P	98
3.5.3	Generating new population vs. re-weighting	101
3.5.4	The marginal likelihood as predictive distribution	102
3.5.5	The eMC ⁴ sampler	103
3.5.6	Evaluation—proof of concept	104
3.5.7	Conclusion	109
4	Ad-hoc and heuristic methods	110
4.1	Available cases analysis	110
4.2	Bound and Collapse—BC	111
4.3	Markov Blanket Predictor—MBP	114
4.3.1	The general idea	115
4.3.2	Approximate predictive distributions	115
4.3.3	Parameter estimation	117
4.3.4	Prediction and missing parents	119
4.3.5	Predictive quality	120
4.3.6	Selecting predictive variables	120
4.3.7	Implementation of MBP	123
4.3.8	Parameter estimation	124
4.3.9	Model learning	126
4.3.10	Conclusion and discussion	128
6.	CONCLUSION	131
	References	133