Bibliography

[1] U. C. Irvine Machine Learning Repository. ftp://ftp.ics.uci.edu/pub/machine-learning-databases/.

[2] Ann Becker and Dan Geiger. A sufficiently fast algorithm for finding close to optimal junction trees. In *UAI 96 Proceedings*, 1996.

[3] C. Berrou and A. Glavieux. Near-optimum error correcting coding and decoding Turbo codes. *IEEE Transactions on Communications*, 44:1261–1271, 1996.

[4] P. Brucker. On the complexity of clustering algorithms. In R. Henn, B. Corte, and W. Oletti, editors, *Optimierung und Operations Research*, Lecture Notes in Economics and Mathematical Systems, pages 44–55. Springer Verlag, 1978.

[5] Peter Cheeseman and John Stutz. *Bayesian classification (AutoClass): Theory and results*, pages 153–180. AAAI Press, 1995.

[6] Jie Cheng, David A. Bell, and Weiru Liu. Learning belief networks from data: an information theory based approach. In *Proceedings of the Sixth ACM International Conference on Information and Knowledge Management*, 1997.

[7] C. K. Chow and C. N. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, IT-14(3):462–467, May 1968.

[8] Gregory F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[9] Gregory F. Cooper and Edward Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[10] Thomas H. Cormen, Charles E. Leiserson, and Ronald R. Rivest. *Introduction to Algorithms*. MIT Press, 1990.

[11] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[12] Robert Cowell. Sampling without replacement in junction trees. Statistical Research Paper 15, City University, London, 1997.

[13] Sanjoy Dasgupta. Learning polytrees. 1998.

[14] A. P. Dawid. Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, 2:25–36, 1992.

[15] Peter Dayan and Richard S. Zemel. Competition and multiple cause models. *Neural Computation*, 7(3), 1995.

[16] Luis M. de Campos and Juan F. Huete. Algorithms for learning decomposable models and chordal graphs. In Dan Geiger and Prakash Pundalik Shenoi, editors, *Proceedings of the 13th Conference on Uncertainty in AI*, pages 46–53. Morgan Kaufmann, 1997.

[17] Carl G. de Marcken. Dependency detection in subquadratic time. 1998.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39:1–38, 1977.

[19] Denise L. Draper and Steve Hanks. Localized partial evaluation of belief networks. In *Proceedings of the 10th Conference on Uncertainty in AI*. Morgan Kaufmann Publishers, 1994.

[20] Michael L. Fredman and Robert Endre Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Jounal of the Association for Computing Machinery*, 34(3):596–615, July 1987.

[21] Brendan J. Frey, Geoffrey E. Hinton, and Peter Dayan. Does the wake-sleep algorithm produce good density estimators? In D. Touretsky, M. Mozer, and M. Hasselmo, editors, *Neural Information Processing Systems*, number 8, pages 661–667. MIT Press, 1996.

[22] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29:131–163, 1997.

[23] Nir Friedman, Moises Goldszmidt, and Tom Lee. Bayesian network classification with continous attributes: Getting the best of both discretization and parametric fitting. In *Proceedings of the International Conference on Machine Learning (ICML)*.

[24] Nir Friedman and Moses Goldszmidt. Building classifiers using Bayesian networks. In *Proceedings of the National Conference on Artificial Intelligence (AAAI 96)*, pages 1277–1284, Menlo Park, CA, 1996. AAAI Press.

[25] H. N. Gabow, Z. Galil, T. Spencer, and Robert Endre Tarjan. Efficient algorithms for finding minimum spanning trees in undirected and directed graphs. *Combinatorica*, 6(2):109–122, 1986.

[26] Robert G. Gallager. *Low-density parity-check codes.* MIT Press, 1963.

[27] Dan Geiger. An entropy-based learning algorithm of bayesian conditional trees. In *Proceedings of the 8th Conference on Uncertainty in AI*, pages 92–97. Morgan Kaufmann Publishers, 1992.

[28] Dan Geiger. Knowledge representation and inference in similarity networks and bayesian multinets. *"Artificial Intelligence"*, 82:45–74, 1996.

[29] Dan Geiger and Christopher Meek. Graphical models and exponential families. In *Proceedings of the 14th Conference on Uncertainty in AI*, pages 156–165. Morgan Kaufmann Publishers, 1998.

[30] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984.

[31] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Chapman & Hall, London, 1996.

[32] David Heckerman. A tutorial on learning Bayesian networks. Technical Report MSR–TR–95–06, Microsoft Research, 1995.

[33] David Heckerman, Dan Geiger, and David M. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.

[34] Reimar Hofmann and Volker Tresp. Nonlinear Markov networks for continuous variables. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Neural Information Processing Systems*, number 10, pages 521–529. MIT Press, 1998.

[35] Tommi S. Jaakkola. *Variational methods for inference and estimation in graphical models*. PhD thesis, Massachusetts Institute of Technology, 1997.

[36] Finn V. Jensen. *An introduction to Bayesian networks*. Springer, 1996.

[37] Finn V. Jensen and Frank Jensen. Optimal junction trees. In *Proceedings of the 10th Conference on Uncertainty in AI*. Morgan Kaufmann Publishers, 1994.

[38] Finn V. Jensen, Steffen L. Lauritzen, and K. G. Olesen. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, 4:269–282, 1990.

[39] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. *Learning and inference in graphical models*, chapter "An introduction to variational methods for graphical models", pages 75–104. 1998.

[40] Uffe Kjærulff. Approximation of Bayesian networks through edge removals. Technical Report Report, Aarlborg University, Department of Mathematics and Computer Science, 94–2007, 1993.

[41] Petri Kontkanen, Petri Myllymaki, and Henry Tirri. Constructing bayesian finite mixture models by the EM algorithm. Technical Report C-1996-9, Univeristy of Helsinky, Department of Computer Science, 1996.

[42] S. Kotz and N. L. Johnson (Eds.). *Encyclopedia of Statistical Sciences*. Wiley, 1982–1989.

[43] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, March 1951.

[44] Steffen L. Lauritzen. *Graphical models*. Clarendon Press, Oxford, 1996.

[45] Steffen L. Lauritzen and Nanni Wermuth. Mixed interaction models. Research report R-84-8, Institute for Electronic Systems, University of Arlborg, Denmark, 1989.

[46] David J. C. MacKay and Radford M. Neal. Near Shannon limit performance of low density parity check codes. *Electronics Letters*, 33:457–458, 1997.

[47] David Madigan and Adrian Rafferty. Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, 89, 1994.

[48] Marina Meilă and Michael I. Jordan. Estimating dependency structure as a hidden variable. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Neural Information Processing Systems*, number 10, pages 584–590. MIT Press, 1998.

[49] Marina Meilă, Michael I. Jordan, and Quaid D. Morris. Estimating dependency structure as a hidden variable. Technical Report AIM–1648,CBCL–165, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1998. (revised version of AIM–1611).

[50] D. Michie, D. J. Spiegelhalter, and C. C. Taylor. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood Publishers, 1994.

[51] Stefano Monti and Gregory F. Cooper. A Bayesian network classfier that combines a finite mixture model and a naive Bayes model. Technical Report ISSP-98-01, University of Pittsburgh, March 1998.

[52] Radford M. Neal. Connectionist learning of belief networks. *Artificial Intelligence*, 56:71–113, 1992.

[53] Hermann Ney, Ute Essen, and Reinhard Kneser. On structuring probabilistic dependences in stochastic language modelling. *Computer Speech and Language*, 8:1–38, 1994.

[54] Michiel O. Noordewier, Geoffrey G. Towell, and Jude W. Shawlik. Training Knowledge-Based Neural Networks to recognize genes in DNA sequences. In Richard P. Lippmann, John E. Moody, and David S. Touretztky, editors, *Advances in Neural Information Processing Systems*, volume 3, pages 530–538. Morgan Kaufmann Publishers, 1991.

[55] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman Publishers, San Mateo, CA, 1988.

[56] Carl E. Rasmussen, Radford M. Neal, Geoffrey E. Hinton, Drew van Camp, Michael Revow, Zoubin Ghahramani, R. Kustra, and Robert Tibshrani. *The DELVE Manual*. http://www.cs.utoronto.ca/ delve, 1996.

[57] Jorma Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Publishing Company, New Jersey, 1989.

[58] Jeff Schlimmer. Mushroom database. U.C. Irvine Machine Learning Repository.

[59] Raffaella Settimi and Jim Q. Smith. On the geometry of Bayesian graphical models with hidden variables. In *Proceedings of the 14th Conference on Uncertainty in AI*, pages 472–479. Morgan Kaufmann Publishers, 1998.

[60] R. Sibson. SLINK: an optimally efficient algorithm for the single-link cluster method. *The Computer Journal*, 16, 1973.

[61] D. J. Spiegelhalter, A. Thomas, N. G. Best, and W. R. Gilks. *BUGS: Bayesian inference Using Gibbs Sampling, Version 3.0*. Medical Research Council Biostatistics Unit, Cambridge, 1994.

[62] Peter Spirtes and Thomas Richardson. A polynomial time algorithm for determining DAG equivalence in the presence of latent variables and selection bias. In *Proceedings of the AISTATS-97 workshop*, pages 489–500, 1997.

[63] Elena Stanghellini and Barbara Vantaggi. On identification of graphical log-linear models with one unobserved variable. Technical Report 99/1, Universitá di Perugia, 1999.

[64] Evan W. Steeg. *Automated motif discovery in protein structure prediction.* PhD thesis, University of Toronto, 1997.

[65] Robert Endre Tarjan. *Data structures and network algorithms.* Society for Industrial and Applied Mathematics, 1983.

[66] Bo Thiesson, Christopher Meek, D. Maxwell Chickering, and David Heckerman. Learning mixtures of Bayes networks. Technical Report MSR–POR–97–30, Microsoft Research, 1997.

[67] Geoffrey Towell and Jude W. Shawlik. Interpretation of artificial neural networks: Mapping Knowledge Based Neural Networks into rules. In John E. Moody, Steve J. Hanson, and Richard P. Lippmann, editors, *Advances in Neural Information Processing Systems*, volume 4, pages 9–17. Morgan Kaufmann Publishers, 1992.

[68] Thomas Verma and Judea Pearl. An algorithm for deciding if a set of observed independencies has a causal explanation. In Didier Dubois, Michael P. Wellman, Bruce D'Ambrosio, and Philippe Smets, editors, *Proceedings of the 8th Conference on Uncertainty in AI*, pages 323–330. Morgan Kaufmann, 1992.

[69] James D. Watson, Nancy H. Hopkins, Jeffrey W. Roberts, Joan Argetsinger Steitz, and Alan M. Weiner. *Molecular Biology of the Gene*, volume I. The Benjamin/Cummings Publishing Company, 4 edition, 1987.

[70] Yair Weiss. Belief propagation and revision in networks with loops. Technical Report AIM–1616,CBCL–155, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1997.

[71] Douglas B. West. *Introduction to Graph Theory.* Prentice Hall, 1996.

[72] Joe Whittaker. *Graphical models in applied multivariate statistics.* John Wiley & Sons, 1990.