

**Eggeling, Ralf; Grosse, Ivo; Koivisto, Mikko**

**Algorithms for learning parsimonious context trees.** (English) Zbl 07073611  
Mach. Learn. 108, No. 6, 879-911 (2019).

Summary: Parsimonious context trees, PCTs, provide a sparse parameterization of conditional probability distributions. They are particularly powerful for modeling context-specific independencies in sequential discrete data. Learning PCTs from data is computationally hard due to the combinatorial explosion of the space of model structures as the number of predictor variables grows. Under the score-and-search paradigm, the fastest algorithm for finding an optimal PCT, prior to the present work, is based on dynamic programming. While the algorithm can handle small instances fast, it becomes infeasible already when there are half a dozen four-state predictor variables. Here, we show that common scoring functions enable the use of new algorithmic ideas, which can significantly expedite the dynamic programming algorithm on typical data. Specifically, we introduce a memoization technique, which exploits regularities within the predictor variables by equating different contexts associated with the same data subset, and a bound-and-prune technique, which exploits regularities within the response variable by pruning parts of the search space based on score upper bounds. On real-world data from recent applications of PCTs within computational biology the ideas are shown to reduce the traversed search space and the computation time by several orders of magnitude in typical cases.

**MSC:**

**68T05** Learning and adaptive systems in artificial intelligence

**Keywords:**

[exact algorithms](#); [structure learning](#); [context-specific independence](#); [branch and bound](#); [sequential data](#)

**Software:**

[JASPAR](#); [Jstacs](#); [UCI-ml](#); [UniProt](#)

**Full Text:** [DOI](#)

**References:**

- [1] Akaike, H., A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19, 716-723, (1974) · [Zbl 0314.62039](#)
- [2] Bacardit, J.; Stout, M.; Hirst, J.; Valencia, A.; Smith, R.; Krasnogor, N., Automated alphabet reduction for protein datasets, *BMC Bioinformatics*, 10, 6, (2009)
- [3] Barash, Y., Elidan, G., Friedman, N., & Kaplan, T. (2003). Modeling dependencies in protein-DNA binding sites. In [\textit{Proceedings of the seventh annual international conference on research in computational molecular biology \(RECOMB\)}](#) (pp 28-37).
- [4] Begleiter, R.; El-Yaniv, R.; Yona, G., On prediction using variable order Markov models, *Journal of Artificial Intelligence Research*, 22, 385-421, (2004) · [Zbl 1148.60306](#)
- [5] Ben-Gal, I.; Shani, A.; Gohr, A.; Grau, J.; Arviv, S.; Shmilovici, A.; Posch, S.; Grosse, I., Identification of transcription factor binding sites with variable-order Bayesian networks, *Bioinformatics*, 21, 2657-2666, (2005)
- [6] Bertsimas, D.; Dunn, J., Optimal classification trees, *Machine Learning*, 106, 1039-1082, (2017) · [Zbl 06841416](#)
- [7] Blanchard, G.; Schäfer, C.; Rozenholc, Y.; Müller, K., Optimal dyadic decision trees, *Machine Learning*, 66, 209-241, (2007)
- [8] Bourguignon, P. Y., & Robelin, D. (2004). Modèles de Markov parcimonieux: sélection de modèle et estimation. In [\textit{Proceedings of the 5e édition des Journées Ouvertes en Biologie, Informatique et Mathématiques \(JOBIM\)}](#).
- [9] Boutillier, C., Friedman, N., Goldszmidt, M., & Koller, D. (1996). Context-specific independence in Bayesian networks. In [\textit{Proceedings of the 12th conference on uncertainty in artificial intelligence \(UAI\)}](#) (pp. 115-123).
- [10] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). [\textit{Classification and regression trees}](#). Belmont: Wadsworth. · [Zbl 0541.62042](#)
- [11] Brocchieri, L.; Karlin, S., Protein length in eukaryotic and prokaryotic proteomes, *Nucleic Acids Research*, 33, 3390-3400, (2005)

- [12] Bühlmann, P.; Wyner, A., Variable length Markov chains, *Annals of Statistics*, 27, 480-513, (1999) · [Zbl 0983.62048](#)
- [13] Buntine, W., Learning classification trees, *Statistics and Computing*, 2, 63-73, (1992)
- [14] Chavira, M., & Darwiche, A. (2005). Compiling Bayesian networks with local structure. In *Proceedings of the 19th international joint conference on artificial intelligence (IJCAI)* (pp. 1306-1312).
- [15] Chickering, D., Heckerman, D., & Meek, C. (1997). A Bayesian approach to learning Bayesian networks with local structure. In *Proceedings of the 13th conference on uncertainty in artificial intelligence (UAI)* (pp. 80-89).
- [16] Chipman, H.; George, E.; McCulloch, R., Bayesian CART model search, *Journal of the American Statistical Association*, 93, 935-948, (1998)
- [17] Campos, C.; Ji, Q., Efficient structure learning of Bayesian networks using constraints, *Journal of Machine Learning Research*, 12, 663-689, (2011) · [Zbl 1280.68226](#)
- [18] Dempster, A.; Laird, N.; Rubin, D., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, 39, 1-38, (1977) · [Zbl 0364.62022](#)
- [19] Eggeling, R. (2018). Disentangling transcription factor binding site complexity. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky688> (epub ahead of print).
- [20] Eggeling, R.; Gohr, A.; Keilwagen, J.; Mohr, M.; Posch, S.; Smith, A.; Grosse, I., On the value of intra-motif dependencies of human insulator protein CTCF, *PLoS ONE*, 9, e85-629, (2014)
- [21] Eggeling, R.; Grosse, I.; Grau, J., InMoDe: Tools for learning and visualizing intra-motif dependencies of DNA binding sites, *Bioinformatics*, 33, 580-582, (2017)
- [22] Eggeling, R., & Koivisto, M. (2016). Pruning rules for learning parsimonious context trees. In *Proceedings of the 32nd conference on uncertainty in artificial intelligence (UAI)* (pp. 152-161).
- [23] Eggeling, R., Koivisto, M., & Grosse, I. (2015a). Dealing with small data: On the generalization of context trees. In *Proceedings of the 32nd international conference on machine learning (ICML)* (pp. 1245-1253).
- [24] Eggeling, R., Roos, T., Myllymäki, P., & Grosse, I. (2014b). Robust learning of inhomogeneous PMMs. In *Proceedings of the 17th international conference on artificial intelligence and statistics (AISTATS)* (pp. 229-237).
- [25] Eggeling, R.; Roos, T.; Myllymäki, P.; Grosse, I., Inferring intra-motif dependencies of DNA binding sites from ChIP-seq data, *BMC Bioinformatics*, 16, 375, (2015)
- [26] Frank, E. (2000). *Pruning decision trees and lists*. Ph.D. Thesis, University of Waikato, Department of Computer Science, Hamilton, New Zealand.
- [27] Fujimaki, R., & Morinaga, S. (2012). Factorized asymptotic Bayesian inference for mixture modeling. In *Proceedings of the 15th international conference on artificial intelligence and statistics (AISTATS)*.
- [28] Grau, J.; Keilwagen, J.; Gohr, A.; Haldemann, B.; Posch, S.; Grosse, I., Jstacs: A Java framework for statistical analysis and classification of biological sequences, *Journal of Machine Learning Research*, 13, 1967-1971, (2012) · [Zbl 06276171](#)
- [29] Heckerman, D.; Geiger, D.; Chickering, D., Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*, 20, 197-243, (1995) · [Zbl 0831.68096](#)
- [30] Hush, D.; Porter, R., Algorithms for optimal dyadic decision trees, *Machine Learning*, 80, 85-107, (2010)
- [31] Jaeger, M.; Nielsen, J.; Silander, T., Learning probabilistic decision graphs, *International Journal of Approximate Reasoning*, 42, 84-100, (2006) · [Zbl 1096.68704](#)
- [32] Kangas, K., Koivisto, M., & Niinimäki, T. (2014). Learning chordal Markov networks by dynamic programming. In *Advances in neural information processing systems (NIPS)* (Vol. 27, pp. 2357-2365).
- [33] Leonardi, F., A generalization of the PST algorithm: Modeling the sparse nature of protein sequences, *Bioinformatics*, 22, 1302-1307, (2006)
- [34] Li, T.; Fan, K.; Wang, J.; Wang, W., Reduction of protein sequence complexity by residue grouping, *Protein Engineering*, 16, 323-330, (2003)
- [35] Lichman, M. (2013). *UCI machine learning repository*. Irvine, CA: University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/ml>. Accessed 8 Oct 2018.
- [36] Lomax, S.; Vadera, S., A survey of cost-sensitive decision tree induction algorithms, *ACM Computing Surveys*, 45, 16:1-16:35, (2013) · [Zbl 1293.68232](#)
- [37] Nielsen, S., The stochastic EM algorithm: Estimation and asymptotic results, *Bernoulli*, 6, 457-489, (2000) · [Zbl 0981.62022](#)
- [38] Oliver, J. (1993). Decision graphs—an extension of decision trees. In *Proceedings of the 4th international workshop on artificial intelligence and statistics (AISTATS)* (pp. 343-350).
- [39] Ordoñez, F.; Toledo, P.; Sanchis, A., Activity recognition using hybrid generative/discriminative models on home environments using binary sensors, *Sensors*, 13, 5460-5477, (2013)
- [40] Orenstein, Y.; Shamir, R., A comparative analysis of transcription factor binding models learned from PBM, HT-SELEX and ChIP data, *Nucleic Acids Research*, 42, e63, (2014)
- [41] Peterson, E.; Kondev, J.; Theriot, J.; Phillips, R., Reduced amino acid alphabets exhibit an improved sensitivity and selectivity in fold assignment, *Bioinformatics*, 25, 1356-1362, (2009)
- [42] Quinlan, JR, Induction of decision trees, *Machine Learning*, 1, 81-106, (1986)
- [43] Rantanen, K., Hyttinen, A., & Järvisalo, M. (2017). Learning chordal Markov networks via branch and bound. In *Advances in neural information processing systems* (NIPS), (Vol. 30, pp. 1845-1855).

- [44] Rissanen, J., A universal data compression system, *IEEE Transactions on Information Theory*, 29, 656-664, (1983) · [Zbl 0521.94010](#)
- [45] Sandelin, A.; Alkema, W.; Engström, P.; Wasserman, W.; Lenhard, B., JASPAR: An open-access database for eukaryotic transcription factor binding profiles, *Nucleic Acids Research*, 32, d91-d94, (2004)
- [46] Schneider, T.; Stephens, R., Sequence logos: A new way to display consensus sequences, *Nucleic Acids Research*, 18, 6097-6100, (1990)
- [47] Schwarz, G., Estimating the dimension of a model, *Annals of Statistics*, 2, 461-464, (1978) · [Zbl 0379.62005](#)
- [48] Seifert, M.; Gohr, A.; Strickert, M.; Grosse, I., Parsimonious higher-order hidden Markov models for improved array-CGH analysis with applications to *Arabidopsis thaliana*, *PLOS Computational Biology*, 8, e1002-286, (2012)
- [49] Shen, Y., Choi, A., & Darwiche, A. (2018). Conditional PSDDs: Modeling and learning with modular knowledge. In *Proceedings of the 33rd national conference on artificial intelligence (AAAI)* (pp. 6433-6442).
- [50] Silander, T., & Myllymäki, P. (2006). A simple approach for finding the globally optimal Bayesian network structure. In *Proceedings of the 22nd annual conference on uncertainty in artificial intelligence (UAI)*.
- [51] Silander, T.; Roos, T.; Myllymäki, P., Learning locally minimax optimal Bayesian networks, *International Journal of Approximate Reasoning*, 51, 544-557, (2010)
- [52] Smith, J.; Anderson, P., Conditional independence and chain event graphs, *Artificial Intelligence*, 172, 42-68, (2008) · [Zbl 1182.68303](#)
- [53] Su, J., & Zhang, H. (2005). Representing conditional independence using decision trees. In *Proceedings of the 20th national conference on artificial intelligence (AAAI)* (pp. 874-879).
- [54] Teyssier, M., & Koller, D. (2005). Ordering-based search: A simple and effective algorithm for learning Bayesian networks. In *Proceedings of the 21st conference on uncertainty in artificial intelligence (UAI)* (pp. 584-590).
- [55] The UniProt Consortium, UniProt: The universal protein knowledgebase, *Nucleic Acids Research*, 45, d158-d169, (2017)
- [56] Tian, J. (2000). A branch-and-bound algorithm for MDL learning in Bayesian networks. In *Proceedings of the 16th conference on uncertainty in artificial intelligence (UAI)* (pp. 580-588).
- [57] Volf, P., & Willems, F. (1994). Context maximizing: Finding MDL decision trees. In *Proceedings of 15th symposium on information theory, Benelux* (pp. 192-200).
- [58] Wilcoxon, F., Individual comparisons by ranking methods, *Biometrics Bulletin*, 1, 80-83, (1945)
- [59] Zhao, X.; Huang, H.; Speed, T., Finding short DNA motifs using permuted Markov models, *Journal of Computational Biology*, 12, 894-906, (2005)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.