Exact Methods for Bayesian Network Structure Learning

Simon de Givry (CR HdR), George Katsirelos (CR)
INRA, MIA Toulouse

1 Context

Bayesian Networks (BNs) are graphical models which allow us to express a joint probability distribution over a set of random variables by exploiting conditional independence between variables to decompose the distribution.

The problem of Bayesian Network Structure Learning consists in finding a Bayesian Network with maximum likelihood given a set of observations over discrete random variables. This problem has several applications including gene regulation networks [ACADG$^+$13], risk analysis [TCRG08] and image processing [LSS05]. Unfortunately it is a computationally challenging problem, as it involves exploring the space of all directed acyclic graphs, which is superexponential in the number of random variables. In terms of computational complexity, it is NP-hard [CHM04], meaning that it is unlikely that it can be solved in polynomial time.

Some recent progress in solving this problem in practice has been achieved by limiting the number of parents of each random variables, that is, by limiting the number of other variables on which it is immediately dependent. Several approaches have proved successful, based on integer programming [BC17], dynamic programming [FY15] and constraint programming [vBH15]. This progress also relies on a range of techniques that improve local inference.

The SaAB team of MIAT has previously worked on this problem as well as its applications on reconstruction of gene regulatory networks [ACADG$^+$13]. A previous thesis explored incomplete methods for this using biological data in a genetical genomic context [VMdG12]. Additionally, the team has extensive experience in the development of practical methods for combinatorial optimization, in particular in the context of cost function networks (CFNs) [HOA$^+$16].

2 Thesis project

The objective of the thesis is to develop a set of techniques for solving the Bayesian Network Structure Learning problem as a Cost Function Network optimization problem. CFNs offer a flexible framework for expressing optimization problems. The TOULBAR2 solver developed by the SaAB team has proved extremely successful as a black box solver. In this project we will go beyond that and develop new optimization techniques, both general and specific for the BNSL problem. We will build on the results of a previous internship in this direction and develop a propagator for the weighted acyclicity constraint. We will incorporate symmetry breaking and dominance constraints developed in the context of a constraint programming approach [vBH15], as well as cut generation methods developed for an integer programming approach [BC17], following our previous work in the same direction [dGK17]. Additionally, we will expand the existing component caching facilities of the TOULBAR2 solver to handle components that do not arise from decomposition. Finally, we will extend the solver to handle implictly defined exponentially larger variable domains, in order to handle BNSL problems with no bound on the arity of the parent sets. In a more general direction, we will augment the branching heuristics of the solver to take into account the information produced by the dual bounds, building on existing work [HSS18] and the results of a previous internship with the team.

The techniques that we develop will be evaluated on both existing datasets as well as real biological data from the SUNRISE project. The implementation will be incorporated into the open source TOULBAR2 C++ solver, with special provisions for handling BNSL problems directly, without forcing the user to perform conversion to CFN manually. The scientific results will be published in top conferences and journals on Artificial Intelligence and Bioinformatics, as is the normal practice of the SaAB team.

3 PhD funding

This 3-year PhD proposal is funded by the French ANR project DEMOGRAPH.
`http://www.lsis.org/demograph`

References

[ACADG+13] David Allouche, Christine Cierco-Ayrolles, Simon De Givry, Gérald Guillermin, Brigitte Mangin, Thomas Schiex, Jimmy Vandel, and Matthieu Vignes. A panel of learning methods for the reconstruction of gene regulatory networks in a systems genetics context. In *Gene Network Inference*, pages 9–31. Springer, 2013.

[BC17] Mark Bartlett and James Cussens. Integer linear programming for the bayesian network structure learning problem. *Artificial Intelligence*, 244:258 – 271, 2017. Combining Constraint Solving with Mining and Learning.

[CHM04] David Maxwell Chickering, David Heckerman, and Christopher Meek. Large-sample learning of bayesian networks is NP-hard. *J. Mach. Learn. Res.*, 5:1287–1330, December 2004.

[dGK17] Simon de Givry and George Katsirelos. Clique cuts in weighted constraint satisfaction. In J. Christopher Beck, editor, *Principles and Practice of Constraint Programming*, pages 97–113, Cham, 2017. Springer International Publishing.

[FY15] Xiannian Fan and Changhe Yuan. An improved lower bound for bayesian network structure learning. In *AAAI Conference on Artificial Intelligence*, 2015.

[HOA+16] Barry Hurley, Barry O'Sullivan, David Allouche, George Katsirelos, Thomas Schiex, Matthias Zytnicki, and Simon de Givry. Multi-language evaluation of exact solvers in graphical model discrete optimization. *Constraints*, 21(3):413–434, Jul 2016.

[HSS18] Stefan Haller, Paul Swoboda, and Bogdan Savchynskyy. Exact map-inference by confining combinatorial search with LP relaxation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018*, 2018.

[LSS05] Jiebo Luo, Andreas E Savakis, and Amit Singhal. A bayesian network-based framework for semantic image understanding. *Pattern recognition*, 38(6):919–934, 2005.

[TCRG08] Paolo Trucco, Enrico Cagno, Fabrizio Ruggeri, and Ottavio Grande. A bayesian belief network modelling of organisational factors in risk analysis: A case study in maritime transportation. *Reliability Engineering & System Safety*, 93(6):845–856, 2008.

[vBH15] Peter van Beek and Hella-Franziska Hoffmann. Machine learning of bayesian networks using constraint programming. In Gilles Pesant, editor, *Principles and Practice of Constraint Programming*, pages 429–445, Cham, 2015. Springer International Publishing.

[VMdG12] J. Vandel, B Mangin, and S. de Givry. New local move operators for bayesian network structure learning. In *Proceedings of Probabilistic Graphical Models*, 2012.