

LEARNING PROBABILISTIC GRAPHICAL MODELS OF GENE NETWORKS AND FAULT NETWORKS

JOSE M. PEÑA
ADIT, IDA, LINKÖPING UNIVERSITY, SE-58183 LINKÖPING, SWEDEN
JOSE.M.PENA@LIU.SE

ABSTRACT. This note is the final report of the project. It consists of an introduction to the project, followed by the results obtained and, finally, a list of the publications produced.

1. INTRODUCTION

One of the most commonly used approaches to model a system is that of using a probability distribution for the task. It has the advantage of having solid foundations and a large body of knowledge exists, which allows advanced reasoning with the model. For instance, one can compute posterior probability distributions conditioned on some observations made. However, this approach does not scale well unless the probability distribution encodes the independences that follow from the system being modeled. For the user to accept such a model, it is crucial that these independences are represented in a visual manner, e.g. by graphs. This is the motivation of an increasingly popular branch of computer science and statistics called (probabilistic) graphical models. It has the advantage that graphs are well studied and many of the operations that are required to reason with the probability distribution can be performed with the aid of the graphs efficiently.

There are two questions that need to be addressed before using a graph to represent the independences in a probability distribution or an arbitrary set of independences (also called independence model) for that matter. First, what type of edges to include in the graph (directed, undirected, bidirected) and, second, attach an interpretation to the (lack of) edges in the graph. When the edges in the graph are all directed / undirected / bidirected, the graph is called a Bayesian network (BN) / Markov network (MN) / covariance graph (CovGs). When the edges in the graph are directed and undirected or directed and bidirected, the graph is called a chain graph (CG). While there is consensus on the semantics of a BN, MN and CovG, there are three different interpretations of CGs in the literature, each with its own advantages and disadvantages. These are known as Lauritzen-Wermuth-Frydenberg (LWF), Andersson-Madigan-Perlman (AMP), and multivariate regression (MVR) interpretation.

BNs are undoubtedly the most popular of all the families of graphical models. However, any of the interpretations of CGs is to be preferred, since CGs are a superset of BNs and thus they are more expressive. However, expressiveness is not all that matters for the widespread use of a family of graphical models. It is equally if not more important that the family allows performing efficient learning and inference. That is, for the full expressiveness of CGs to be deployed in practice, one also needs efficient algorithms for reasoning with them, and efficient algorithms for learning them from data. This was the main objective of this project: To develop the machinery needed for CGs to take off. In our humble opinion, we think that we have contributed decisively to it.

We thank CENIIT for giving us the possibility of carrying out this project. Their funding was primarily used to cover part of my salary. This allowed me to devote more time to the supervision of my PhD student Dag Sonntag, co-author of many of the publications produced in this project. It also allowed me to collaborate with researchers at the University's hospital, which resulted in some publications in the medical field. Finally, it allowed me to start a new line of research on algorithmic trading together with my recently recruited PhD student Marcus Bendtsen.

2. RESULTS

In this section, we list the main results achieved in the project. The results have been split in four areas. The publications corresponding to each of the areas are listed in the next section.

2.1. Chain graphs: Expressiveness.

- We have compared the expressiveness of the three different interpretations of CGs and characterize when an independence model represented by one interpretation can also be represented by the others.
- We have solved some formal issues that may have hindered the widespread use of AMP CGs such as the lack of an expression to factorize a probability distribution according to the CG, the lack of an efficient method to learn the parameters of the distribution, and the lack of an efficient inference method to reason with the distribution.
- We have shown that LWF and AMP CGs do not represent arbitrary independence models by showing that every model represented corresponds to the result of marginalization and conditioning in a causal model. For MVR CGs, this result is trivial.
- We have shown that LWF and AMP CGs do not represent arbitrary independence models by showing that every model represented corresponds to some discrete and continuous probability distribution. For MVR CGs, this result is trivial.
- We have quantified how much more expressive CGs are compared to BNs and MNs.
- We have combined AMP and MVR CGs into a new family of graphical models that we call marginal AMP CGs (MAMP CGs). We have shown that there are independence models that can be represented by the new family but not by AMP or MVR CGs.

2.2. Chain graphs: Learning.

- We have proposed efficient learning algorithms for AMP and MVR CGs under the assumption that the learning data has been sampled from a probability distribution that is faithful to an AMP or MVR CG, respectively. That is, the sampled distribution has all and only the independences represented by the CG.
- We have proposed an efficient learning algorithm for LWF CGs under the assumption that the learning data has been sampled from a probability distribution that satisfies the so-called composition property. This assumption is much weaker than the faithfulness assumption. For instance, every regular Gaussian distribution satisfies the composition property assumption but not necessarily the faithfulness assumption. Unfortunately, such an efficient algorithm cannot be developed for AMP and MVR CGs since they do not satisfy the so-called Meek's conjecture. We proved Meek's conjecture for LWF CGs, which was a crucial step to develop the mentioned learning algorithm.
- We have developed an algorithm for learning LWF CGs without any assumption about the sampled probability distribution. The new algorithm is based on answer set programming and performs an exhaustive search over the whole space of CGs efficiently. However, scalability is an issue. This is the price to pay for the lack of assumptions.
- We have proposed an efficient learning algorithm for MAMP CGs under the faithfulness assumption.

2.3. Chain graphs: Biomedical applications.

- We have written a review chapter on graphical models for representing gene networks. We have argued why CGs are to be preferred.
- We have developed a system based on BNs to predict the risk of a cardiac event. Unlike other similar systems, ours is patient-specific in the sense that the system recommends the next measurements to collect so as to optimize the prediction for the patient being treated. This is a collaboration with researchers at the University's hospital.

2.4. Other topics.

- We have shown how to read dependencies off different families of graphical models, whose original goal is to represent independences. In other words, we have characterized the dependencies that follow from the independences not represented by the model.

- We have shown how to obtain a consensus graphical model from the individual models provided by different experts or obtained from different learning datasets. The goal is to obtain a single model that only represents independences represented by all the individual models, and as many of them as possible.
- Together with my PhD student Marcus Bendtsen, we have started investigating the benefits of graphical models for algorithmic trading, i.e. the detection of optimal points in time to buy and sell stocks. For this purpose, we have introduced a new family of graphical models called gated BNs that allow us to represent dynamic systems consisting of a recurrent non-deterministic sequence of phases or regimes.

3. LIST OF PUBLICATIONS

3.1. Chain graphs: Expressiveness.

- Peña, J. M. (2015). Factorization, Inference and Parameter Learning in Discrete AMP Chain Graphs. In Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty (ECSQARU 2015) Lecture Notes in Artificial Intelligence, to appear.
- Peña, J. M. (2015). Every LWF and AMP Chain Graph Originates from a Set of Causal Models. In Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty (ECSQARU 2015) Lecture Notes in Artificial Intelligence, to appear.
- Sonntag, D., Peña, J. M. and Gmez-Olmedo, M. (2015). Approximate Counting of Graphical Models Via MCMC Revisited. *International Journal of Intelligent Systems*, 30 (3), 384420.
- Sonntag, D. and Peña, J. M. (2015). On Expressiveness of the Chain Graph Interpretations. Submitted.
- Sonntag, D. and Peña, J. M. (2015). Chain Graph Interpretations and their Relations Revisited. *International Journal of Approximate Reasoning*, accepted.
- Peña, J. M. (2014). Marginal AMP Chain Graphs. *International Journal of Approximate Reasoning*, 55 (5), 1185-1206.
- Peña, J. M. (2013). Error AMP Chain Graphs. In Proceedings of the 12th Scandinavian Conference on Artificial Intelligence (SCAI 2013) - *Frontiers in Artificial Intelligence and Applications* 257, 215-224.
- Sonntag, D. and Peña, J. M. (2013). Chain Graph Interpretations and their Relations. In Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty (ECSQARU 2013) Lecture Notes in Artificial Intelligence 7958, 510-521.
- Peña, J. M. (2013). Approximate Counting of Graphical Models Via MCMC Revisited. In Proceedings of the 15th Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2013) - *Lecture Notes in Artificial Intelligence* 8109, 383-392.
- Peña, J. M. (2011). Faithfulness in Chain Graphs: The Gaussian Case. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011), 588-599.
- Peña, J. M. (2009). Faithfulness in Chain Graphs: The Discrete Case. *International Journal of Approximate Reasoning*, 50 (8), 1306-1313.

3.2. Chain graphs: Learning.

- Sonntag, D., Järvisalo, M., Peña, J. M. and Hyttinen, A. (2015). Learning Optimal Chain Graphs with Answer Set Programming. In Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence (UAI 2015), to appear.
- Peña, J. M. and Gomez-Olmedo, M. (2015). Learning Marginal AMP Chain Graphs under Faithfulness Revisited. Submitted.
- Peña, J. M. (2014). Learning Marginal AMP Chain Graphs under Faithfulness. In Proceedings of the 7th European Workshop on Probabilistic Graphical Models (PGM 2014) Lecture Notes in Artificial Intelligence 8754, 382-395.

- Peña, J. M., Sonntag, D. and Nielsen, J. D. (2014). An Inclusion Optimal Algorithm for Chain Graph Structure Learning. In Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS 2014), 778-786.
- Peña, J. M. (2014). Learning AMP Chain Graphs and some Marginal Models Thereof under Faithfulness. *International Journal of Approximate Reasoning*, 55 (4), 1011-1021.
- Peña, J. M. (2012). Learning AMP Chain Graphs under Faithfulness. In Proceedings of the 6th European Workshop on Probabilistic Graphical Models (PGM 2012), 251-258.
- Sonntag, D. and Peña, J. M. (2012). Learning Multivariate Regression Chain Graphs under Faithfulness. In Proceedings of the 6th European Workshop on Probabilistic Graphical Models (PGM 2012), 299-306.

3.3. Chain graphs: Biomedical applications.

- al Sam, S., Åhre, M., Peña, J. M., Länne, T., Östgren, C. J., Nyström, F. and Cedersund, G. Patient-Specific Combinations of Biomarkers for Calculating the Risk of Cardiac Events. In preparation.
- Sonntag, D. and Peña, J. M. (2015). Chain Graphs and Gene Networks. In *Foundations of Biomedical Knowledge Representation*. Springer, accepted.

3.4. Other topics.

- Keivani, O. and Peña, J. M. (2015). Uni- and Multi-Dimensional Clustering Via Bayesian Networks. In *Unsupervised Learning Algorithms*. Springer, accepted.
- Bendtsen, M. and Peña, J. M. (2015). Gated Bayesian Networks for Algorithmic Trading. Submitted.
- Bendtsen, M. and Peña, J. M. (2014). Learning Gated Bayesian Networks for Algorithmic Trading. In Proceedings of the 7th European Workshop on Probabilistic Graphical Models (PGM 2014) *Lecture Notes in Artificial Intelligence* 8754, 49-64.
- Bendtsen, M. and Peña, J. M. (2013). Gated Bayesian Networks. In Proceedings of the 12th Scandinavian Conference on Artificial Intelligence (SCAI 2013) - *Frontiers in Artificial Intelligence and Applications* 257, 35-44.
- Etminani, K., Naghibzadeh, M. and Peña, J. M. (2013). DemocraticOP: A Democratic Way of Aggregating Bayesian Network Parameters. *International Journal of Approximate Reasoning*, 54 (5), 602614.
- Peña, J. M. (2013). Reading Dependencies from Covariance Graphs. *International Journal of Approximate Reasoning*, 54 (1), 216-227.
- Peña, J. M. (2011). Finding Consensus Bayesian Network Structures. *Journal of Artificial Intelligence Research*, 42, 661-687.
- Peña, J. M. (2010). Reading Dependencies from Polytree-Like Bayesian Networks Revisited. In Proceedings of the 5th European Workshop on Probabilistic Graphical Models (PGM 2010), 225-232.