Efficient Query Answering in Nonparametric Probabilistic Graphical Models*

Mattis Hartwig

University of Lübeck, Institute of Information Systems
{hartwig}@ifis.uni-luebeck.de

Abstract. This paper describes the focus of the dissertation to combine probabilistic graphical models with nonparametric approaches. We discuss related work, develop a research framework and present first results of one and a half year of research.

Keywords: Probabilistic Graphical Models · Nonparametrics · query answering.

1 Introduction

The probabilistic graphical model language is frequently used in medical diagnosis, bio-statistics, ecology, maintenance, etc. and results in probabilistic graphical models (PGMs) [22] [13]. In a PGM, variables are modeled as nodes and the relationships between the variables as connections between the nodes [11]. The strength of the influence (modeled as arrows), is contained in a conditional probability table. For each state of the influencing variable, the influenced variable has a different probability distribution. One famous PGM based application called Pathfinder assists practitioners with diagnosis of lymph-node diseases [8].

Of course PGMs also have downsides. Two basic versions of PGMs allow for either only discrete variables or for only Gaussian distributed variables which is a strong limitation when representing real-world phenomena [11]. These limitations that restrict the potential outcomes to certain families/types of functions (e.g. on Gaussian distributions) are called parametric assumptions. This means that PGMs alone do not fulfill the four characteristics mentioned above. Models that do not have these strong assumptions are called nonparametric. Two promising nonparametric concepts are infinite mixture models and Gaussian processes (GPs). Mixture model based methods combine as many individual models as needed to get a combined more complex model with less limitations. GP based methods are mostly used in time series modeling and represent a distribution over functions by making probabilistic assumptions about similarity of observed and unobserved data points [16, 18]. Often these nonparametric approaches bring other downsides like higher computational costs.

* Supported by Studienstiftung des deutschen Volkes

Consequently, the dissertation attempts to combine PGMs with mixture models and GPs. Additionally, new algorithms to use the models will be developed. If possible a real-world application from healthcare is used as an evaluation for the developed approaches. This results in 4 research questions for the dissertation:

RQ 1: How can PGMs and nonparametric approaches be combined to increase the expressive power of the resulting model?
RQ 2: How can existing query answering algorithms both be transferred to new graphical models with higher expressive power?
a) approximate vs. exact
b) grounded vs. lifted
RQ 3: How can new models and query answering algorithms support real-world application in healthcare?

2 Related Work

In this section, the current state of the research along the dimensions of the research framework is reviewed. Table 1 contains the main literature by subtopic for each dimension.

Table 1. Most relevant literature by research dimension and subtopic

Dimension	**Subtopic**	**Most relevant literature**
Models	Probabilistic graphical models	[11, 14]
	Mixture models	[2, 6]
	Gaussian processes	[16, 1, 17]
Query answering	Exact Inference	[12, 19]
	Approximate Inference	[3, 9]
	Lifting	[20, 4, 5]

Bayesian networks have been already combined with nonparametric mixture models to get nonparametric Bayesian networks where the number of components can basically grow dependent on the data fed into the model [10, 7]. Another nonparametric model that can be combined with PGMs is the GP model. [16] has triggered a lot of research in the area of GPs. A good intuition for a GP is a distribution over functions. This means instead of learning one single function or a single distribution, GPs allow to learn a Bayesian distribution over functions.

GPs use kernel functions (also known as covariance functions) to describe how strongly correlated different data points are (usually depending on their distance). If one point is observed it influences the distribution in the other points. GPs are often used to describe time series models. Therefore they can be related with Dynamic PGMs which also describe developments along a time

series. [21] formulated in his PhD Thesis: “All of the typical linear time series models such as autoregressive, moving average, autoregressive moving average, and the Kalman Filter are equivalent to a GP with the appropriate covariance matrix.” The challenge is to find the appropriate kernel function that results in a desired covariance matrix structure. [17] have explored the kernel function for a specific Near Constant Acceleration Kalman filter and [16] has described the kernel function for a simple markov chain.
The existing links between PGMs and GPs are limited to one random variable. In another context multi-output GPs have been already used [1]. The dissertation will continue the work on bringing PGMs and GPs together while also including multi-output situations.

For exact query answering in PGMs we build upon the well-know junction tree algorithm developed by Lauritzen et al. [12] and combine it with the new model types. As much as possible of the query answering should also be transferred to the lifted case. In lifting isomorphic random variables are grouped together and algorithms try to only use representatives for the groups to speed up query answering [15]. Braun and Möller [4] have already lifted the junction tree algorithm.

3 Dissertation Plan

The research questions can be structured into a research frame work. Figure 1 shows that each question is one dimension of the framework.

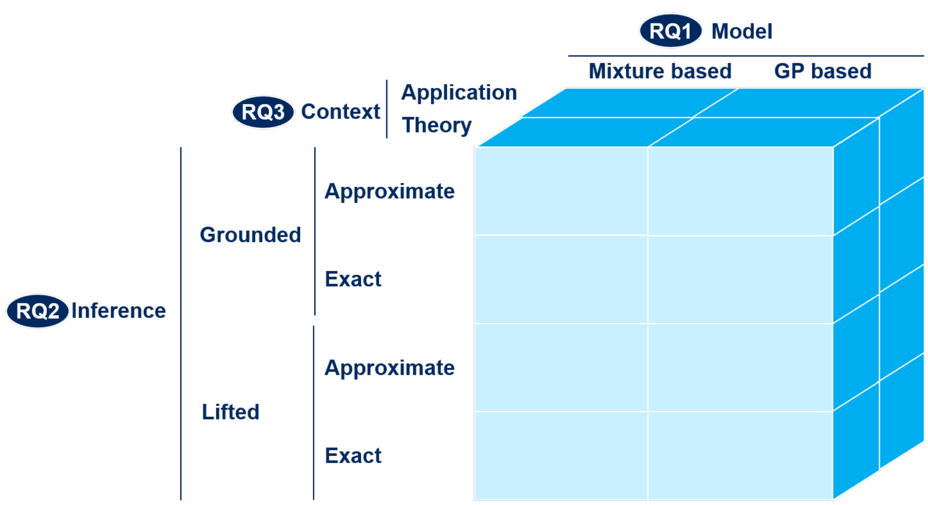

Fig. 1. Translation of research questions into general research framework

In the related work, we already described research that has already been performed in the framework but the aim of the dissertation is to full this cube as dense as possible.

4 First Results

In the first one and a half years of the dissertation first research results have been generated:

- **Efficient Query Answering in Complex Gaussian Mixture models:** The paper describes an approach for approximate inference in Gaussian Mixture models when the number of components gets very large. Published in Proceedings of the ICBK 2019, 2019, IEEE, p.81-86.
- **Constructing Gaussian Processes for Probabilistic Graphical Models:** The paper describes the conversion of two types of PGMs into GPs and develops the according kernel function. Published in Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS-20), 2020.
- **Lifted Query Answering in Gaussian Bayesian Networks:** The paper describes a lifted query answering algorithm for Gaussian Bayesian Networks. Accepted for publishing at the Conference for Probabilistic Graphical Models 2020.
- **Constructing Gaussian Processes for Dynamic Gaussian Bayesian Networks:** The paper describes a more general kernel function allowing the construction of GP for multidimensional Dynamic Gaussian Bayesian Networks with arbitrary inter-time connections. Under Review.

5 Outlook

Up until now we developed first results in most of the theoretical bricks of our framework. To answer research question 1 only one further generalization of PGMs into GPs is missing. The open pieces to answer research question 1 are mainly to look into lifted query answering for Gaussian Processes and to look into generalizing the lifted query answering for Gaussian Bayesian Networks to also handle hybrid networks. For research question 3 we have established a connection to university hospital to detail out a real-world application to evaluate our theoretical work.

References

1. Alvarez, M.A., Rosasco, L., Lawrence, N.D., Others: Kernels for vector-valued functions: A review. Foundations and Trends®in Machine Learning **4**(3), 195–266 (2012)
2. Bishop, C.M.: Pattern recognition and machine learning. Springer (2006)

3. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational Inference: A Review for Statisticians (apr 2017). https://doi.org/10.1080/01621459.2017.1285773
4. Braun, T., Möller, R.: Lifted junction tree algorithm. In: Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz). pp. 30–42. Springer (2016)
5. Gehrke, M., Braun, T., Möller, R.: Lifted dynamic junction tree algorithm. In: International Conference on Conceptual Structures. pp. 55–69. Springer (2018)
6. Gormley, I.C., Frühwirth-Schnatter, S.: Mixtures of Experts Models. arXiv preprint arXiv:1806.08200 (2018)
7. Hanea, A., Napoles, O.M., Ababei, D.: Non-parametric Bayesian networks: Improving theory and reviewing applications. Reliability Engineering & System Safety **144**, 265–284 (2015). https://doi.org/https://doi.org/10.1016/j.ress.2015.07.027, http://www.sciencedirect.com/science/article/pii/S0951832015002331
8. Heckerman, D.E., Horvitz, E.J., Nathwani, B.N.: Toward normative expert systems: Part i the pathfinder project. Methods of information in medicine **31**(02), 90–105 (1992)
9. Hrycej, T.: Gibbs sampling in Bayesian networks. Artificial Intelligence **46**(3), 351–363 (1990)
10. Ickstadt, K., Bornkamp, B., Grzegorczyk, M., Wieczorek, J., Sheriff, M.R., Grecco, H.E., Zamir, E.: Nonparametric Bayesian Networks. Bayesian Stat **9**, 283 (2011)
11. Koller, D., Friedman, N., Bach, F.: Probabilistic graphical models: principles and techniques. MIT press (2009)
12. Lauritzen, S.L., Jensen, F.: Stable local computation with conditional Gaussian distributions. Statistics and Computing **11**(2), 191–203 (apr 2001). https://doi.org/10.1023/A:1008935617754
13. McCann, R.K., Marcot, B.G., Ellis, R.: Bayesian belief networks: applications in ecology and natural resource management. Canadian Journal of Forest Research **36**(12), 3053–3062 (2006)
14. Murphy, K.P.: Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. thesis (2002)
15. Poole, D.: First-order probabilistic inference. In: IJCAI. vol. 3, pp. 985–991 (2003)
16. Rasmussen, C.E.: Gaussian processes for machine learning. MIT Press (2006)
17. Reece, S., Roberts, S.: An introduction to Gaussian processes for the Kalman filter expert. In: 2010 13th International Conference on Information Fusion. pp. 1–9. IEEE (2010)
18. Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., Aigrain, S.: Gaussian processes for time-series modelling. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **371**(1984), 20110550 (2013)
19. Shachter, R.D., Kenley, C.R.: Gaussian Influence Diagrams. Management Science **35**(5), 527–550 (1989). https://doi.org/10.1287/mnsc.35.5.527
20. Taghipour, N., Fierens, D., Davis, J., Blockeel, H.: Lifted Variable Elimination: Decoupling the Operators from the Constraint Language. Journal of Artificial Intelligence Research **47**, 393–439 (jul 2013). https://doi.org/10.1613/jair.3793
21. Turner, R.D.: Gaussian processes for state space models and change point detection. Ph.D. thesis, University of Cambridge (2012)
22. Weber, P., Medina-Oliva, G., Simon, C., Iung, B.: Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Engineering Applications of Artificial Intelligence **25**(4), 671–682 (2012)