Research Statement: Past Accomplishments and Future Plans

Martin Trapp July 23, 2020

My research interests focus on probabilistic machine learning approaches that exploit structural properties, e.g. symmetries or decomposability, to allow efficient and exact computations while maintaining the capability to model complex relationships in data. Much of my work has centered around the intersection of sum-product networks and Bayesian approaches, such as i) learning sum-product networks by formulating the learning task as Bayesian inference over the structure and parameters of the network [11] and ii) exploiting the structural properties of sum-product networks for efficient and exact posterior inference in deep mixtures of Gaussian processes [5]. In addition to my research on tractable probabilistic models, I started investigating probabilistic programming systems and joined the open-source project Turing.jl[1] as a core team member in mid-2018. The goal of the Turing.jl project is to spur research on probabilistic inference and modelling.

Throughout my PhD studies, I have established various close collaborations with leading researcher who share the vision to push modern probabilistic machine learning by exploiting the use of tractable probabilistic inference, including: Zoubin Ghahramani (University of Cambridge and Uber AI), Kristian Kersting (Darmstadt University of Technology), Guy Van den Broeck (UCLA), and Robert Peharz (Eindhoven University of Technology).

Structural properties of modern machine learning models (and their inductive biases) are often key to solve complex modelling tasks effectively, e.g. convolutional neural networks exploit spatial correlations in data to compactly encode and efficiently learn predictors in image domains. In contrast, sum-product networks exploit independence assumptions in subsets of the data to guarantee that global marginalisation tasks simplify to local and tractable marginalisation. Moreover, there is an increasing awareness of the importance of the structural properties of the model architecture used for probabilistic machine learning, e.g. variational autoencoders can be understood as an efficient encoding of infinite mixture models. Thus, illustrating the importance to identify and subsequently exploit structural properties for efficient computations in probabilistic machine learning.

A key challenge in deep tractable probabilistic models, such as sum-product networks, probabilistic sentential decision diagrams and other probabilistic circuits, is the definition of a structure, and therefore choosing an inductive bias of the network, that is suitable for the task at hand. In our recent work [11], we have shown that sum-product network structures can be learned in a fully Bayesian manner, by decomposing the problem by i) defining a so-called computational graph and ii) performing posterior inference over discrete encodings of sub-graphs (of the computational graph) that fulfill certain structural requirements. Thus, our approach poses the structure learning problem as inference over structure encodings, and therefore entails a natural Occam's razor effect and is robust against missing values. The resulting approach is not only conceptually very appealing but also opens the door to many exciting research directions.

[1] The Turing.jl project is organised by Hong Ge and Zoubin Ghahramani from the University of Cambridge. Details about the project can be found on `https://turing.ml`.

However, due to the discrete nature of the inference problem, performing approximate posterior inference over structure encodings can be inefficient and results in poor mixing. Fortunately, many discrete inference problems can be reformulated, e.g. using continuous relaxations [13], to allow more efficient inference through gradient-based learning, e.g. [12]. Exploiting those recent advancements can lead to natural strategies for gradient-based structure and parameter learning in sum-product networks and related model classes. Thus, enabling us to learn the most suitable inductive bias for a given dataset, while maintaining structural properties that guarantee efficient and exact computations.

Moreover, even though we illustrated in [11] that a Bayesian formulation of sum-product networks can be used to learn infinite mixtures over sum-product networks, e.g. for streaming data tasks, exploring infinitely deep and wide sum-product networks learned through parallelisable inference strategies is still an open problem. Fortunately, recent work has shown that completely random measures, which comprise many Bayesian nonparametric models, can be decomposed into two independent sub-measures, the first one being a finite (deterministic) measure and the latter having Levy measure [14]. Exploiting this property has led to exciting advances on inference in infinite mixture models [1] and exploring the decomposability of completely random measures could also lead to parallelisable inference strategies for infinitely wide and deep network structures.

Last but not least, our decomposition of the structure learning problem by defining a computational graph and by learning a discrete encoding, can be understood as finding a binary mask that highlights a suitable dub-graph. This interpretation has some interesting implications, as it enables a more efficient representation of large ensembles of sum-product networks by exploiting recent work on binary matrix factorisations [10] and low-rank approximations for ensembles of deep neural networks [6].

As demonstrated by recent works, exploiting structural properties of probabilistic programs and models can often lead to techniques more suited for efficient probabilistic inference and computations, e.g. [9, 4, 7]. Some recent examples include: i) our paper on Deep Structured Mixtures of Gaussian Processes [5] in which we exploit a nested representation of naive-local experts for efficient and exact posterior inference, and ii) our work on Einsum networks [2] in which we reformulated the likelihood function of layered sum-product networks in terms of Einstein summations – allowing efficient and GPU-accelerated computations. Thus, indicating that leveraging structural properties and model transformations for efficient computations are promising directions.

One key aspect of sum-product networks and other probabilistic circuits is the inherent assumption that the network polynomial of a probability distribution can be efficiently represented by sharing computations and parameters. In fact, sum-product networks can be understood as a compact representation of an exponentially large mixture distribution with combinatorial many mixture components – combinatorial in the latent states of the network – in which network parameters decompose into shared parameters. Campbell, Huggins, How, and Broderick [8] showed that learning nonparametric models using truncated representations exhibits a very slow decay of the truncation error in case of heavy-tailed nonparametric priors. Thus, making it notoriously difficult to learn many common heavy-tailed nonparametric models, such as topic models. Utilising the compact representation of sum-product networks to encode heavy-tailed nonparametric models could, therefore, help in learning such models more effectively.

Last but not least, it is vital to investigate the limitations of recent tractable models and explore alternative formulations. Specifically, it is unclear which model formulations allow efficient and exact computations, while obtaining a maximum in representation capacity. Recently, we have started investigating an approach that exploits symmetries in the data distribution, while allowing exact and efficient integration [3]. However, our current analysis focuses on a very strict sub-class of flow models and an in-depth theoretical investigation is urgently needed.

References

[1] Avinava Dubey, Michael M. Zhang, Eric P. Xing, and Sinead A. Williamson. "Distributed, partially collapsed MCMC for Bayesian nonparametrics". In: *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2020.

[2] Robert Peharz, Steven Lang, Antonio Vergari, Karl Stelzner, Alejandro Molina, Martin Trapp, Guy Van den Broeck, Kristian Kersting, and Zoubin Ghahramani. "Einsum Networks: Fast and Scalable Learning of Tractable Probabilistic Circuits". In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2020.

[3] Tomas Pevny, Vasek Smidl, Martin Trapp, Ondrej Polacek, and Tomas Oberhuber. "Sum-Product-Transform Networks: Exploiting Symmetries using Invertible Transformations". In: *Proceedings of the International Conference on Probabilistic Graphical Models (PGM)*. 2020.

[4] Tom Rainforth, Adam Golinski, Frank Wood, and Sheheryar Zaidi. "Target–Aware Bayesian Inference: How to Beat Optimal Conventional Estimators". In: *Journal of Machine Learning Research* 21.88 (2020), pp. 1–54.

[5] Martin Trapp, Robert Peharz, Franz Pernkopf, and Carl Edward Rasmussen. "Deep Structured Mixtures of Gaussian Processes". In: *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2020.

[6] Yeming Wen, Dustin Tran, and Jimmy Ba. "BatchEnsemble: an Alternative Approach to Efficient Ensemble and Lifelong Learning". In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2020.

[7] Yuan Zhou, Hongseok Yang, Yee Whye Teh, and Tom Rainforth. "Divide, Conquer, and Combine: a New Inference Strategy for Probabilistic Programs with Stochastic Support". In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2020.

[8] Trevor Campbell, Jonathan H. Huggins, Jonathan P. How, and Tamara Broderick. "Truncated random measures". In: *Bernoulli* 25.2 (May 2019), pp. 1256–1288.

[9] Samuel Kolb, Pedro Zuidberg Dos Martires, and Luc De Raedt. "How to Exploit Structure while Solving Weighted Model Integration Problems". In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*. 2019, p. 262.

[10] Ravi Kumar, Rina Panigrahy, Ali Rahimi, and David P. Woodruff. "Faster Algorithms for Binary Matrix Factorization". In: *Proceedings of the International Conference on Machine Learning (ICML)*. Proceedings of Machine Learning Research. 2019, pp. 3551–3559.

[11] Martin Trapp, Robert Peharz, Hong Ge, Franz Pernkopf, and Zoubin Ghahramani. "Bayesian Learning of Sum-Product Networks". In: *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*. 2019, pp. 6344–6355.

[12] Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. "DAGs with NO TEARS: Continuous Optimization for Structure Learning". In: *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*. 2018, pp. 9492–9503.

[13] Chris J. Maddison, Andriy Mnih, and Yee Whye Teh. "The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables". In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2017.

[14] Tamara Broderick, Ashia C. Wilson, and Michael I. Jordan. "Posteriors, conjugacy, and exponential fam-ilies for completely random measures". In: *Bernoulli* 24.4B (2016), pp. 3181–3221.