Uncertainty Quantification: Classical and Quantum Approaches for Limited Labelled-Datasets

Soronzonbold Otgonbaatar, Dieter Kranzlmüller
German Aerospace Center (DLR) Oberpfaffenhofen, Ludwig-Maximilians-Universität München

Abstract

Deep learning (DL) models are extensively used to analyze small- and large-scale datasets due to their scalability and their computational efficiency compared with conventional statistical approaches such as Bayesian analysis. However, they are not capable of learning informative information from small-scale datasets and explaining their predictions; namely, their outputs, given small-scale datasets as input, are not trustworthy and reliable which can be measured by using uncertainty quantification. In fact, DL models are often considered as uninterpretable black boxes with unknown uncertainties, and they even suffer learning on small-scale datasets. In contrast, Bayesian analysis is a gold standard technique for uncertainty quantification in order to obtain trustworthy and reliable predictions generated by models fitted small-scale datasets (observations) due to its high computational cost on processing large-scale datasets. Hence, DL models integrated with Bayesian analysis, that is, Bayesian Neural Networks (BNNs), are slowly gaining a great interest, since they can be trained on both small-and large-scale datasets and allow make their outputs interpretable by yielding trustworthy and reliable uncertainties. However, BNN inference on large-scale datasets persists high computational cost even on supercomputers, and commonly used methodologies to overcome this challenge are Monte Carlo Markov Chain (MCMC) and variational inference (VI) approaches. Moreover, the VI approach, returning approximate samples, can be scaled on big datasets in contrast to the exact sampling MCMC. Therefore, this study assesses and examines quantum VI paradigm for processing BNN inference to improve its sampling power. More importantly, the quantum VI method promises quantum advantage over its classical counterpart, since it can be executed on near- and long-term quantum computers, representing classically intractable probability distributions but hard-to-sample it on a conventional computer in the context of computational complexity hierarchy conjectures.

Classical Bayesian Neural network

Classical Bayesian Neural Networks, for short, Bayesian Neural Networks (CBNNs), are referred to as stochastic Deep Neural networks (DNNs) trained using Bayesian analysis on datasets [1,2]. See below Figure.

- **CBNNs** generate probability distributions of predictions and weights.
- **CBNNs** are natural data-efficient and inherently interpretable models thanks to their respective uncertainties, that is, uncertainties in predictions and weights.
- **CBNNs** are trained approximately via a classical Variational Inference (VI) technique for finding a probability distribution P.
- **DNNs** considered as uninterpretable black-box models require big labeled datasets,
- **DNNs** are needed to be trained and tested on sub-datasets including training, test, and validation sets, while one does not need to divide datasets into training, test, and validation sets for training CBNNs.

$$p(F_{\boldsymbol\theta}|\mathcal{S}) = \frac{p(\mathcal{S}|F_{\boldsymbol\theta})p(\boldsymbol\theta)}{p(\mathcal{S})} \quad \longleftrightarrow \quad p(\boldsymbol\theta|\mathcal{S}) = \frac{p(\mathcal{S}|\boldsymbol\theta)p(\boldsymbol\theta)}{p(\mathcal{S})}$$

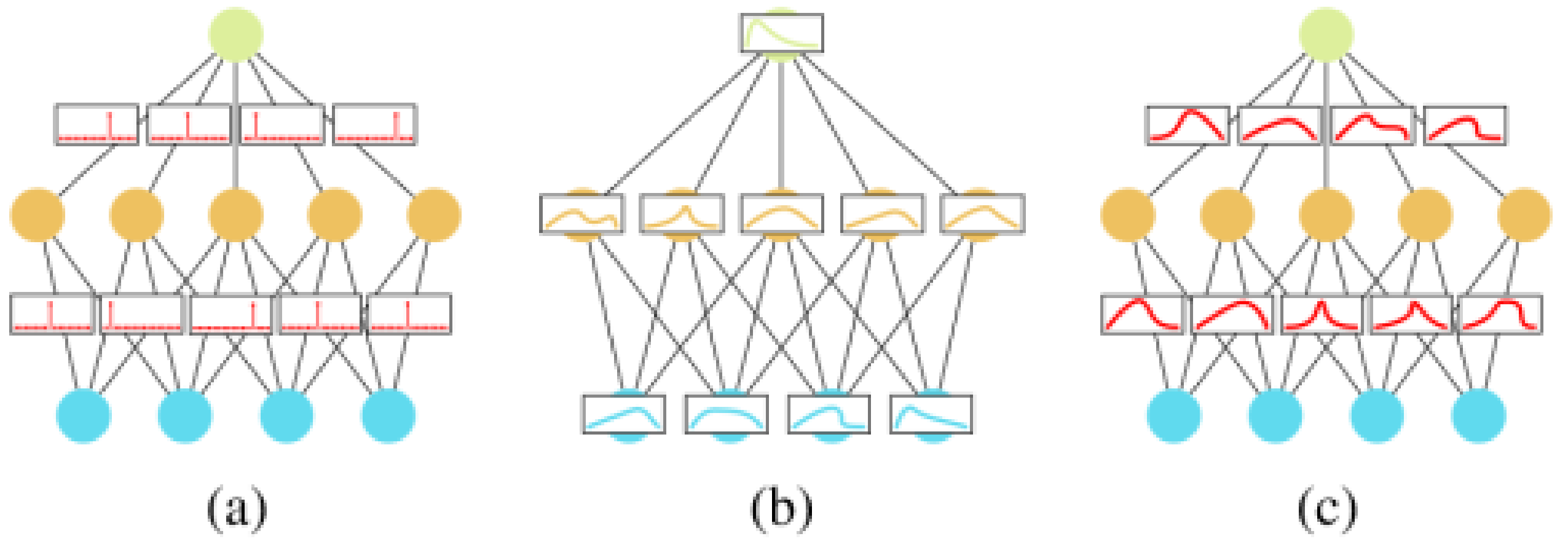

Figure: DNNs versus BNNs: (a) DNNs generate point estimates of predictions based on point weights given training, test, and validation sets, and (b) BNNs generate probability distributions of predictions with (c) a probability distribution over weights given the available dataset, either small or big datasets. Namely, we do not need to divide datasets into training, test, and validation sets thanks to uncertainty information provided by BNNs. Image credit: Jospin [3].

$$\begin{aligned} ELBO(p(\mathcal{S},\boldsymbol\theta)||q_c(\boldsymbol\theta;\boldsymbol\lambda)) &= \mathbb{E}_{q_c(\boldsymbol\theta;\boldsymbol\lambda)}\left[\log(p(\mathcal{S}|\boldsymbol\theta)p(\boldsymbol\theta))\right] - \mathbb{E}_{q_c(\boldsymbol\theta;\boldsymbol\lambda)}\left[\log q_c(\boldsymbol\theta;\boldsymbol\lambda)\right] \\ &= \mathbb{E}_{q_c(\boldsymbol\theta;\boldsymbol\lambda)}\left[\log\frac{p(\mathcal{S},\boldsymbol\theta)}{q_c(\boldsymbol\theta;\boldsymbol\lambda)}\right] = \mathcal{L}(\boldsymbol\lambda,\boldsymbol\theta;\mathcal{S}). \end{aligned}$$

Quantum Bayesian Neural network

Quantum Bayesian Neural Networks (QBNNs) are BNNs boosted by quantum algorithms which are designed to solve efficiently some hard computational problems on quantum computers. Moreover, they promise to generate solutions to a class of computational problems much faster than conventional computing resources, and quantum computers (i.e., quantum circuits) are even able to represent classically intractable probability distributions due to their inherently probabilistic nature and non-classical correlation property, that is, quantum circuits with large entanglement [4-7].

- **QBNNs** make the classical VI better on generating good approximate samples.
- **QBNNs** represent a probability distribution by a parametrized quantum circuit, that is, an Instantaneous Quantum Polynomial (IQP) Circuit.
- **QBNNs** can not be simulated on a conventional computer due to a computational hierarchy conjecture.
- **QBNNs**, hence, exhibit quantum advantage over CBNNs.

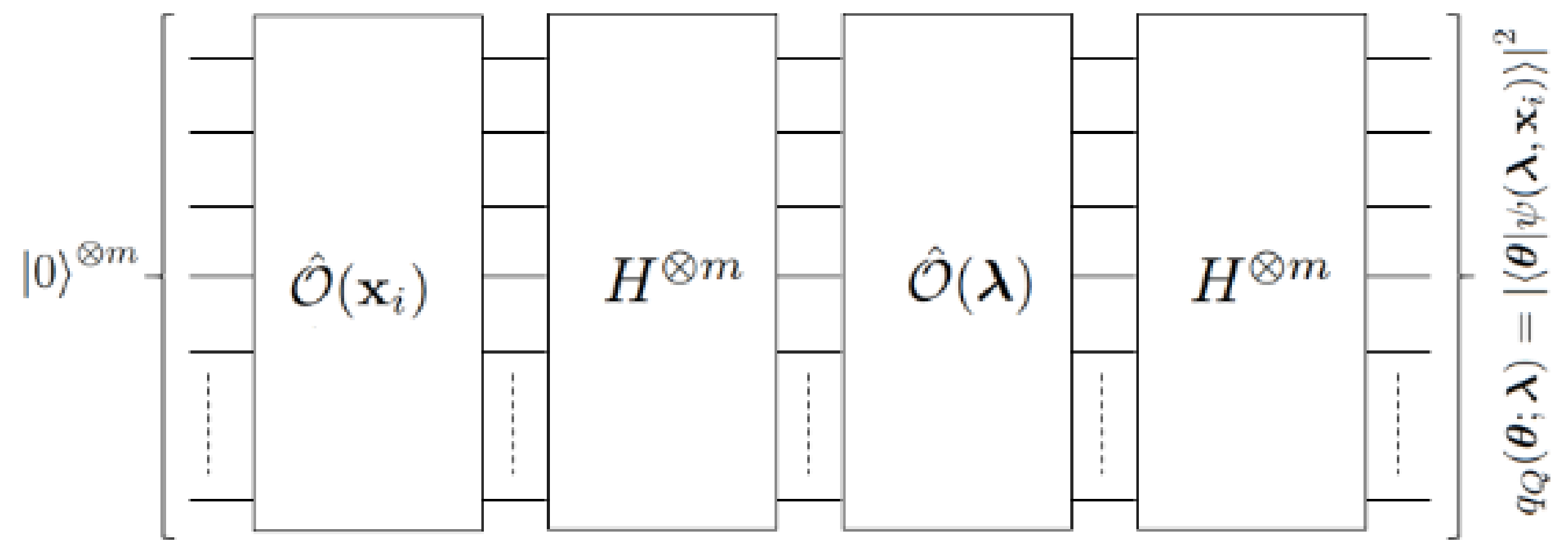

Figure: An Instantaneous Quantum Polynomial Circuit with data encoding quantum gate according to the computational hierarchy conjecture. Namely, if this quantum circuit can be sampled by a classical computer then the computational hierarchy collapses to the so-called third level.

$$ELBO(p(\mathcal{S},\boldsymbol\theta)||q_Q(\boldsymbol\theta;\boldsymbol\lambda)) = \mathbb{E}_{q_Q(\boldsymbol\theta;\boldsymbol\lambda)}\left[\log\frac{p(\mathcal{S},\boldsymbol\theta)}{q_Q(\boldsymbol\theta;\boldsymbol\lambda)}\right] = \mathcal{L}_Q(\boldsymbol\lambda,\boldsymbol\theta;\mathcal{S})$$

Conclusion

Deep Learning (DL) models are extensively applied to process big datasets due to the powerful computing machines like GPU tensor cores and availability of benchmark labeled datasets. They are often considered as uninterpretable black-box models with dubious uncertainties: their outputs are not trustworthy and reliable estimates for making high stake decisions involving EO datasets. As opposed to DL models, classical Bayesian statistical approaches are inherently interpretable models generating trustworthy and reliable predictions with error/uncertainty estimates but there is the challenge that they do not scale well as the size of datasets increases or computationally expensive. This challenge can be tackled by combining the best of both DL model and Bayesian analysis, that is, Bayesian Neural Networks (CBNNs); namely, DL models scale well on increasing the size of benchmark labeled-datasets, while Bayesian approaches generate the trustworthy and reliable predictions with their confidence level. However, CBNNs are still computationally expensive due to their intractable posterior distributions. To weaken CBNNs, variational inference (VI) paradigm approximates the intractable posterior by a tractable variational distribution function by optimizing the ELBO metric. Hence, CBNNs become scalable interpretable models as the size of benchmark label-datasets increases. More importantly, they generalize well on small-scale datasets compared with DL models. There persists, however, the imperfection that the tractable variational distribution returns approximate samples for uncertainty quantification - not exact samples.

To estimate their uncertainties with high precision, we propose a quantum VI instead of its classical counterpart. For quantum VI paradigm, we replace a classical variational distribution function by its quantum version represented by an IQP circuit. According to computational complexity theoretic conjectures, the IQP circuit, that is, superconducting- and photonic-technology IQP circuits, can not be sampled on a conventional computer. This fact proves that quantum variational inference exhibits so-called quantum advantage over its classical counterpart. The quantum variational distribution approximates the intractable posterior better and generates more superior samples for uncertainty quantification than ones generated by the classical variational distribution - closer to exact samples. In particular, the IQP circuit can be executed on superconducting- and photonic technology machines integrated with a classical HPC workflow; HPC+QCs paradigm. The classical part selects informative features from limited labeled-images and performs the dimensionality-reduction on them depending on NISQ machine (3-5 years) or FTQ machine (15 years). The larger and more error-free the qubits, the less the classical resource usage for pre-processing practical datasets.

References

[1] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518):859–877, apr 2017. doi: 10.1080/01621459.2017.1285773.
[2] Y. Gal, P. Koumoutsakos, F. Lanusse, G. Louppe, and C. Papadimitriou. Bayesian uncertainty quantification for machine-learned models in physics. Nature Reviews Physics, 4(9):573–577, Sep 2022. ISSN 2522-5820. doi: 10.1038/s42254-022-00498-4.
[3] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun. Hands-on bayesian neural networks—a tutorial for deep learning users. IEEE Computational Intelligence Magazine, 17(2):29–48, may 2022. doi: 10.1109/mci.2022.3155327
[4] S. Otgonbaatar and M. Datcu. Classification of remote sensing images with parameterized quantum gates. IEEE Geoscience and Remote Sensing Letters, pages 1–5, 2021b. doi: 10.1109/LGRS.2021.3108014.
[5] M. J. Bremner, R. Jozsa, and D. J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 467(2126):459–472, aug 2010. doi: 10.1098/rspa.2010.0301
[6] M. Benedetti, B. Coyle, M. Fiorentini, M. Lubasch, and M. Rosenkranz. Variational inference with a quantum computer. Phys. Rev. Appl., 16:044057, Oct 2021. doi: 10.1103/PhysRevApplied.16.044057.
[7] N. Berner, V. Fortuin, and J. Landman. Quantum bayesian neural networks, 2021, ArXiv: 2107.09599

REMOTE SENSING TECHNOLOGY INSTITUTE

“Toward Quantum Advantage in High Energy Physics”, Max Planck Institute of Quantum Optics (MPQ)