Quantum-Assisted Optimization for the Automation of Accounting

Positioning of the thesis

Sampling and inference are essential parts of a machine learning algorithm. Whenever the program makes a decision, handles predictions and uncertainty or gives solutions to a problem, it is a form of sampling or infering data from a vast set of data based on past data and / or knowledge. When a program *samples* data, it effectively selects portions of the available data or estimates probabilities for events, while *Inference* is the process where the program applies its learned model to make predictions or draw conclusions about new, unseen data. Both ensure that machine learning models are practical, scalable, and effective in solving real-world problems.

But with the emerging problem sizes and state space dimensions (i.e. the complexity of data), both processes turn out to be increasingly hard with the computational cost of exploring the space growing exponentially with the number of variables sometimes even to the NP-hard classes.

A possible path to reduce the computational cost and complexity of these problems could be with quantum mechanics and the use of tensor methods. While tensor decomposition techniques are widely used in machine learning, especially in high-dimensional spaces, to reduce complexity and improve sampling and inference, quantum mechanics offers unique computational capabilities that can make sampling and inference faster and more efficient, particularly for problems that are hard or even NP-hard for classical computers.

Quantum computers can represent and explore many states simultaneously using superposition, making it easier to sample from high-dimensional probability distributions, and with algorithms like Grover's search can amplify the probability of finding a desired outcome, speeding up rare-event sampling or reducing sampling bias. At the same time tensor networks are used to efficiently simulate quantum systems, enabling better understanding and design of quantum algorithms for sampling and inference, sometimes even just to approximate quantum behavior on classical systems, offering some of the benefits of quantum computing without requiring actual quantum hardware.

Objectives and Plan

With these algorithmic approaches posed at the first part of this short roadmap, it would be the main goal of the thesis to leverage the strengths and possibilities posed by them and to provide quantum based tensor network procedure for sampling real data (knowledge bases built on data with logical connectives) that can provide a speedup over purely classical algorithms currently in widespread use.

As a first step it would be needed to represent our datasets and knowledge bases (we can put the two under one hat for now) as Markov Logic Networks (MLNs) as they combine logical and probabilistic approaches through encoding the logical connectives as factors to the graphical model, with further factors used for implementing hard and soft constraints [1][2]. These networks are defined through a partition function and an energy function:

$$P(X) = \frac{1}{Z} exp\left(\sum_i w_i \phi_i(X)\right) \tag{1}$$

They can also be generalized to exponential families. Exponential families are parametrized energy base representations of probability distributions, where each coordinate is defined by a base measure and a set of sufficient statistics over our logical formulas of the knowledge base [3][4]:

$$P(F|\eta) = h(F) exp(\eta^T T(F) - A(\eta)) \tag{2}$$

- F: A logical formula or set of formulas in the KB
- T(F): Sufficient statistics encoding logical structure or truth values.
- η: Natural parameters representing weights or importance of features
- h(F): Base measure
- A(η): Log-partition function for normalization

These representations open up opportunities for sampling with energy based models (where although classical procedures are well defined, the Hamiltonian formalism makes quantum mechanics a powerful tool for evaluating such systems). Secondly it makes rejection sampling of complex distributions easier as the selection of a proposal distribution and the acceptance rate is well defined by the representation of probability distributions within exponential families.

So these representations could be of use when we would like to build a sampling algorithm, for example a quantum Gibbs sampling with energy based models.

After we have a formalism for encoding the datasets as MLNs we could also seen the issue with the curse of dimensionality mentioned before; the computational cost of exploring the space growing exponentially with the number of variables. So by representing these models as tensors, this issue could be mitigated by small distributed representations of the model [5][6], so we could utilize gate based quantum calculations and answer probabilistic queries with contractions.

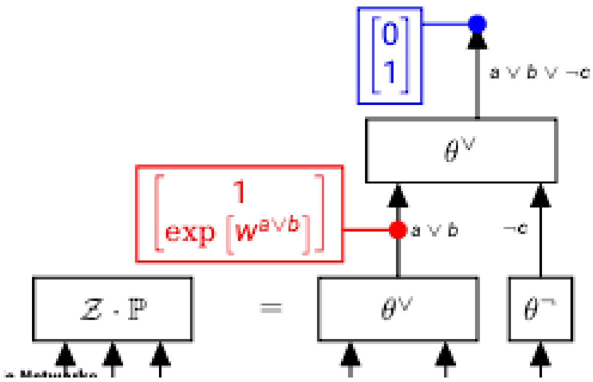

Figure 1: Representing MLNs as Tensor networks

The semantics of complex formulas can be stored by a set of semantics of the used logical connectives as tensors, with the boolean states represented as one-hot encodings (qubits) [7]. Now these tensor networks built on the logic of our knowledge base could be represented as a quantum circuit, where each variable is an input qubit (state with the weights encoded as phases) and all the logical connectives are unitary operations (one/two qubit gates – tensors). A simple sampling procedure could be the contraction over shared indices (the utilization of the quantum circuit) and the resulting values afterwards (the measurement of the resulting states) are our samples.

One last point where we can "win" even more with the utilization of quantum mechanics is the usage of a Quantum Amplitude Amplification algorithm, which in the end could help us amplify the outcome of the lower probability states.

So as a bulletin of the ideas/goals:

- Represent knowledge bases and data as MLNs
- Reshape as Tensor networks
- Map them to quantum circuits with:
  - each variable is an input qubit (state with the weights encoded as phases)
  - logical connectives as unitary operations (one/two qubit gates – tensors)
- perform approximate inference procedures for sampling, like: Quantum Gibbs sampling [8], Quantum rejection sampling [9]
- amplify the probability of finding low probability variables/states with Quantum Amplitude Amplification
- show a measure of speedup (e.g. in the big O notation)

References

1. Bach, S. H., Broecheler, M., Huang, B. & Getoor, L. *Hinge-Loss Markov Random Fields and Probabilistic Soft Logic* 2017. arXiv: 1505.04406 [cs.LG]. https://arxiv.org/abs/1505.04406.
2. Richardson, M. & Domingos, P. Markov logic networks. *Machine Learning* **62,** 107–136. ISSN: 1573-0565. https://doi.org/10.1007/s10994-006-5833-1 (2006).
3. Geiger, D., Heckerman, D., King, H. & Meek, C. Stratified exponential families: Graphical models and model selection. *The Annals of Statistics* **29,** 505 –529. https://doi.org/10.1214/aos/1009210550 (2001).
4. Wainwright, M. J. & Jordan, M. I. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning* **1,** 1–305. ISSN: 1935-8237. http://dx.doi.org/10.1561/2200000001 (2008).
5. Huggins, W., Patil, P., Mitchell, B., Whaley, K. B. & Stoudenmire, E. M. Towards quantum machine learning with tensor networks. *Quantum Science and Technology* **4,** 024001. ISSN: 2058-9565. http://dx.doi.org/10.1088/2058-9565/aaea94 (Jan. 2019).
6. Rieser, H.-M., Köster, F. & Raulf, A. P. Tensor networks for quantum machine learning. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **479.** ISSN: 1471-2946. http://dx.doi.org/10.1098/rspa.2023.0218 (July 2023).
7. Goessmann, A. The tensor network approach to neuro-symbolic AI. *ENEXA proj.* (Nov. 2024).
8. Wittek, P. & Gogolin, C. Quantum Enhanced Inference in Markov Logic Networks. *Scientific Reports* **7.** ISSN: 2045-2322. http://dx.doi.org/10.1038/srep45672 (Apr. 2017).
9. Low, G. H., Yoder, T. J. & Chuang, I. L. Quantum inference on Bayesian networks. *Physical Review A* **89.** ISSN: 1094-1622. http://dx.doi.org/10.1103/PhysRevA.89.062315 (June 2014).