Quantum Techniques in Machine Learning

Dipartimento di Informatica
Università di Verona, Italy

November 6–8, 2017
Contents

Abstracts

Marco Loog — Surrogate Losses in Classical Machine Learning (Invited Talk) . . . 11
Minh Ha Quang — Covariance matrices and covariance operators in machine
learning and pattern recognition: A geometrical framework (Invited Talk) . . . . . 12
Jonathan Romero, Jonathan Olson, Alan Aspuru-Guzik — Quantum au-
toencoders for efficient compression of quantum data . . . . . . . . . . . . . . . . . . 13
Iris Agresti, Niko Viggianiello, Fulvio Flamini, Nicolò Spagnolo, Andrea
Crespi, Roberto Osellame, Nathan Wiebe, and Fabio Sciarrino — Pattern
recognition techniques for Boson Sampling Validation . . . . . . . . . . . . . . . . . . 15
J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M.
Petruzzella, A. Laing, J. G. Rarity, J. L. O’Brien, and M. G. Thompson — Quantum Hamiltonian learning using Bayesian inference on a quantum photonic
simulator . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
Luca Innocenti, Leonardo Banchi, Alessandro Ferraro, Sougato Bose and
Mauro Paternostro — Supervised learning of time independent Hamiltonians for
gate design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
Davide Venturelli — Challenges to Practical End-to-end Implementation of Quan-
tum Optimization Approaches for Combinatorial problems (Invited Talk) . . . . . 22
K. Imafuku, M. Hioki, T. Katashita, S. Kawabata, H. Koike, M. Maezawa,
T. Nakagawa, Y. Oiwa, and T. Sekigawa — Annealing Computation with
Adaptor Mechanism and its Application to Property-Verification of Neural Network
Systems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
Simon E. Nigg, Niels Niels Lörch, Rakesh P. Tiwari — Robust quantum
optimizer with full connectivity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
William Cruz-Santos, Salvador E. Venegas-Andraca and Marco Lanzagorta
— Adiabatic quantum optimization applied to the stereo matching problem . . . . 26
Alejandro Perdomo Ortiz — Opportunities and Challenges for Quantum-Assisted
Machine Learning in Near-Term Quantum Computers . . . . . . . . . . . . . . . . . . 27
Christopher J. Turner, Konstantinos Meichanetzidis, Zlatko Papić, and
Jiannis K. Pachos — Distinguishing free and interacting as pattern recognition . . 28
Konstantinos Meichanetzidis, Christopher J. Turner, Ashk Farjami, Zlatko
Papić, and Jiannis K. Pachos — Free-fermion descriptions of parafermion chains
and string-net models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
Ashk Farjami — Identifying Free Particle Correlations in Topologically Ordered
Systems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
Ivan Glasser, Nicola Pancotti, Moritz August, Ivan D. Rodriguez, and J.
Ignacio Cirac — Neural Networks Quantum States, String-Bond States and chiral
topological states . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
Raban Iten, Roger Colbeck, and Matthias Christandl — Quantum Circuits
for Quantum Channels . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
Seth Lloyd — Prospects in Quantum Machine Learning (Invited Talk) . . . . . 38
Jiannis Pachos — Knots, Computation and Quantum Physics (Invited Talk) . . 39
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Towards Quantum Machine Learning with Tensor Networks</td>
<td>40</td>
</tr>
<tr>
<td>Quantum Walk Neural Networks</td>
<td>43</td>
</tr>
<tr>
<td>Quantum machine learning for quantum anomaly detection</td>
<td>45</td>
</tr>
<tr>
<td>A quantum causal discovery algorithm</td>
<td>47</td>
</tr>
<tr>
<td>Progress in Quantum Reinforcement Learning (Invited Talk)</td>
<td>49</td>
</tr>
<tr>
<td>Quantum Neural Networks: A Hamiltonian Complexity Approach</td>
<td>50</td>
</tr>
<tr>
<td>Measuring Entanglement Negativity with Neural Network Estimators</td>
<td>54</td>
</tr>
<tr>
<td>Quantum state engineering using one-dimensional discrete-time quantum walks</td>
<td>56</td>
</tr>
<tr>
<td>Quantum Entanglement Simulators Inspired by Tensor Network</td>
<td>53</td>
</tr>
<tr>
<td>Measuring Entanglement Negativity with Neural Network Estimators</td>
<td>54</td>
</tr>
<tr>
<td>Quantum state engineering using one-dimensional discrete-time quantum walks</td>
<td>56</td>
</tr>
<tr>
<td>Quantum Error Correction with Recurrent Neural Networks</td>
<td>63</td>
</tr>
<tr>
<td>Quantum learning of coherent states</td>
<td>65</td>
</tr>
<tr>
<td>Inductive supervised quantum learning</td>
<td>69</td>
</tr>
<tr>
<td>Adversarial Domain Adaptation for Identifying Quantum Phase Transitions</td>
<td>71</td>
</tr>
<tr>
<td>Quantum Change Point Identification</td>
<td>72</td>
</tr>
<tr>
<td>A Quantum-inspired version of the Nearest Mean Classifier</td>
<td>73</td>
</tr>
<tr>
<td>Quantum Machine Learning with Small-Scale Devices</td>
<td>75</td>
</tr>
<tr>
<td>Demonstration of Envariance and Parity Learning on the IBM 16 Qubit Processor</td>
<td>76</td>
</tr>
<tr>
<td>Learning Noise in Quantum Information Processors</td>
<td>77</td>
</tr>
<tr>
<td>Quantum enhanced neural network architecture evaluation</td>
<td>81</td>
</tr>
<tr>
<td>Learning and Controlling a Spin Environment through Quantum Measurements</td>
<td>83</td>
</tr>
<tr>
<td>Extractable Work-Based Criteria and Geometric Quantum Speed</td>
<td>84</td>
</tr>
<tr>
<td>Programming Quantum Computers: functional languages and new perspectives</td>
<td>86</td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>Modelling Multipartite Entanglement in Quantum Protocols using Evolving Entangled Hypergraphs</td>
<td>Linda Anticoli and Masoud Gharahi Ghahi</td>
</tr>
<tr>
<td>Quantum Machine Group Learning</td>
<td>Carlos Tavares</td>
</tr>
</tbody>
</table>
Surrogate Losses in Classical Machine Learning  
(Invited Talk)

Marco Loog
TU Delft, Nederlands

Because of the computational complexity of various optimization problems in machine learning, one often resorts to the minimization of a so-called surrogate loss that turns the inherently problematic objective function into a convex one. Supervised classification is a basic setting in which such surrogate losses are widely used. The discrepancy between the loss of ultimate interest and the loss optimized can cause unexpected, even counterintuitive, behaviour. This poses challenges with respect to the development of learning theories. I exemplify the latter issue through the problem of semi-supervised learning — in particular the question in what way it can lead to improved performance — and sketch two possible routes to its resolution. Finally, some related open problems in active learning, transfer learning, etc. are briefly discussed and I will speculate on the benefits of actually being able to optimize the loss of interest.
Covariance matrices and covariance operators in machine learning and pattern recognition: A geometrical framework
(Invited Talk)

Minh Ha Quang
Italian Institute of Technology, Genoa, Italy

Symmetric positive definite (SPD) matrices, in particular covariance matrices, play important roles in many areas of mathematics and statistics, with numerous applications in various different fields, including machine learning, brain imaging, and computer vision. The set of SPD matrices is not a subspace of Euclidean space and consequently algorithms utilizing only the Euclidean metric tend to be suboptimal in practice. A lot of recent research has therefore focused on exploiting the intrinsic geometrical structures of SPD matrices, in particular the view of this set as a Riemannian manifold. In this talk, we will present a survey of some of the recent developments in the generalization of finite-dimensional covariance matrices to infinite-dimensional covariance operators via kernel methods, along with the corresponding geometrical structures. This direction exploits the power of kernel methods from machine learning in the framework of Riemannian geometry, both mathematically and algorithmically. The theoretical formulation will be illustrated with applications in computer vision, which demonstrate both the power of kernel covariance operators as well as of the algorithms based on their intrinsic geometry.
Quantum autoencoders for efficient compression of quantum data
Jonathan Romero, Jonathan Olson, Alan Aspuru-Guzik
Department of Chemistry and Chemical Biology, Harvard University, Mallinckrodt Laboratory, Cambridge, MA 02138, USA

Classical autoencoders are neural networks that can learn efficient low dimensional representations of data in higher dimensional space. The task of an autoencoder is, given an input \( x \), to map \( x \) to a lower dimensional point \( y \) such that \( x \) can likely be recovered from \( y \). The structure of the underlying autoencoder network can be chosen to represent the data on a smaller dimension, effectively compressing the input.

Inspired by this idea, we introduced the model of a quantum autoencoder to perform similar tasks on quantum data [1]. We show application of the quantum autoencoder to quantum simulation, and discuss potential applications to quantum communication and other information theoretic tasks.

A quantum autoencoder consists of a parameterized quantum circuit that is trained to compress a particular dataset of quantum states. While traditional compression algorithms may require computational resources exponential in the number of qubits, the quantum autoencoder is designed to be able to find a good encoding in polynomial time. Given a set of states, \( \{ |\Psi_i\rangle \}_{i=1,...,N} \), and a choice of state, \( |\alpha\rangle \), the autoencoder finds a parameterized unitary \( U(\vec{p}) \), able to factorize the states as the tensor product:

\[
U(\vec{p})|\Psi_i\rangle = |\alpha\rangle \otimes |\Psi_i^C\rangle \quad \forall \ i \in \{1, ..., N\}
\]

The parameters of the quantum autoencoder, \( \vec{p} \), are optimized using a classical optimization algorithm, as shown in Figure 1. We present an example of a simple programmable circuit that can be trained as an efficient autoencoder. We apply our model in the context of quantum simulation to compress ground states of the Hubbard model and molecular Hamiltonians and in the context of quantum communication to compress information sources. We also explore potential implementations of this model in existing quantum hardware.

Figure 1. Schematic representation of the hybrid scheme employed to train quantum autoencoders. The input state is prepared by application of \( W_i \) and subsequently compressed by the application of \( U^{\vec{p}} \) and measurement of the top qubits. These steps are reversed by appending the reference step and applying the decomposition unitary and the inverse step preparation. Fidelities are estimated as the probability of measuring the initial step at the end of the circuit. Alternatively, fidelities can be estimated as the probability of measuring the reference state after compression. Results for all the states in the training set are collected and the sum of the squared overlaps is maximized using a classical optimization algorithm. The process is repeated until the values of the parameters and the cost function converge.
To illustrate the use of a quantum autoencoder, we simulated the training on a set of fermionic wavefunctions. The compression schemes obtained through the autoencoder procedure could be employed to reduce the requirements of quantum memory, if the wavefunction needed to be stored. It also could save quantum resources for the simulation of systems with similar symmetries.

![Figure 2](image.png)

**Figure 2.** (left) Potential energy surface for the hydrogen molecule using an STO-6G basis set. The solid blue line represents the exact potential. The ground states at the red dots were used as the training set for the quantum autoencoder. The ground states at the blue dots were used for testing. (right) A plot of the cost function versus the number of cost function evaluations during the training process. This corresponds to compression of the wavefunction of H₂ from 4 to 2 qubits using a training set of six ground states. We compared the L-BFGS-B and Basin-Hopping algorithms for optimization.

We point out that the maximum rate of lossless compression achievable with a quantum autoencoder is predetermined by the size of the subspace spanning the training set. Consequently, a given set of states might only admit a small or null compression rate. For instance, consider a fermionic system with 8 fermionic modes and 4 particles, such as a half-filled 4-sites Hubbard model or the H₄ molecule in a minimal basis set studied in [1]. Based solely on the constraints in the number of particles, these 8-qubits systems could be compressed to \(\log_2(\binom{8}{4}) \approx 7\) qubits. In general, we expect fermionic systems where the number of fermionic modes is considerably larger than the number of particles to be good candidates for compression.

Additionally, the model with some variations may have other applications to known protocols, such as compression for quantum communication, error-correcting circuits, state preparation, and estimation of entropy and channel capacities. Quantum autoencoders could be used as a quantum classifier or as a subroutine in other machine learning algorithms.

**References**

Pattern recognition techniques for Boson Sampling Validation

Iris Agresti1, Niko Viggianiello1, Fulvio Flamini1, Nicolò Spagnolo1, Andrea Crespi2,3, Roberto Osellame2,3, Nathan Wiebe4, Fabio Sciarrino1

1 Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy
2 Istituto di Fotonica e Nanotecnologie, Consiglio Nazionale delle Ricerche (IFN-CNR), Piazza Leonardo da Vinci, 32, I-20133 Milano, Italy
3 Dipartimento di Fisica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, I-20133 Milano, Italy
4 Station Q Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA, United States

1 Boson Sampling and its validation

Boson Sampling is a computational model, proposed for the first time in 2010 by S. Aaronson and A. Arkhipov [1]. The flurry of interest that their work has attracted is consequence of the strong theoretical evidence that the two authors provided about the computational hardnes of this task. Indeed, the existence of an efficient classical algorithm to solve Boson Sampling would imply the collapse of the polynomial hierarchy to the third level [1]. This computational complexity of Boson Sampling is on the other hand counterbalanced by an apparent ease of implementation through quantum hardwares, such as photonic platforms. This makes Boson Sampling a most interesting candidate to give evidence of the so-called quantum supremacy, that is the regime in which a quantum computer outperforms its classical counterpart.

The Boson Sampling problem consists in sampling from the output probability distribution obtained after the evolution of \(N\) identical, i.e. indistinguishable, bosons through a \(m\)-mode linear transformation. However, even if we managed to build a Boson Sampler, with \(N\) and \(m\) large enough to show a quantum speedup, we couldn’t prove its correct functioning. The reasons are mainly three: (i) for large \(N\) and \(m\) we can’t evaluate the whole probability distribution through classical computation, so we cannot rely on any theoretical model; (ii) with our finite number of measurements we can not evaluate the distribution we are sampling from; (iii) even if we could perform an infinite number of measurements in order to perfectly approximate the distribution we are sampling from, we could not find whether it were correct or not, since the Boson Sampling distribution is also hard to verify. This opens the path for a branch of physics called Boson Sampling validation, which rules out alternative plausible hypotheses. One of the most relevant is given when distinguishable particles evolve through the unitary transformation, making classical rather than quantum interference occur.

2 Our validation protocol

While many validation protocols have been proposed, much of them require the calculation of permanents (a very hard computational problem), which means they are not scalable for \(N\) and \(m\). Therefore, in our work we deal with a slightly different problem: given a trusted Boson Sampler, we want to efficiently validate another one. Indeed, we require to rely only on experimental data.

To this aim, we adopt the following approach (see also [2]), see Fig.1: (i) we draw an experimental sample from the trusted device and one from the device we want to validate and we consider the Fock representation of the sampled bosonic states as coordinates (input and output states are described by a mode occupation list \(\{ s_1, ..., s_m \}\), where \(s_i\) is the number of bosons in the \(i\)-th mode); (ii) we sort the events drawn from the trusted device in clusters; (iii)
we keep the cluster structure fixed and we put data belonging to the second sample in it, assigning each event to
the cluster whose center is the nearest (we adopt the Euclidean distance); (iv) we perform a compatibility test ($\chi^2$
 test). If the two samples are recognized as compatible, the second device is validated, otherwise not.

Figure 1: **Validation Protocol.** The four steps which constitute our validation protocol: (1.) draw two data
samples, one from the trusted device and one from the device we want to validate; (2.) Build a cluster structure on
the trustworthy sample; (3.) Put the data belonging to the other sample inside the already built cluster structure;
(4.) Perform a compatibility test.

The novelty of our work lies in the second step, where we adopted unsupervised machine learning methods to
build our cluster structure.

Indeed in our work we examined different techniques of clustering, like $K$-means clustering and Hierarchical clus-
tering, considering also different initializations and variations for each technique and we tested the performances
of our protocol in ruling out the alternative hypothesis of distinguishable-particle inputs, on experimental samples
belonging to the scenario of $N=3$, $m=13$. For example in Fig.2 the purple dots show the success percentage our
test in distinguishing incompatible samples, while varying the sample size.

Besides the very high success percentages, one of the most striking feature of our protocol is that, once trained,
the algorithm is still effective for much greater Hilbert spaces and in ruling out alternative scenarios different to
the one for which it was trained. In particular, we tested our optimized test parameters on numerical samples
belonging to scenarios up to $N=7$ and $m=70$, i.e. 7 orders of magnitude larger than the first scenario. As alternative
hypotheses we considered also the Uniform sampler and the Mean-Field sampler, the latter being a scenario firstly
proposed by [3], where particles simulate bosonic interference properties, but whose probability distribution is an
easy computational task.

Figure 2: **Success percentages in discriminating incompatible samples.** Success percentages of our protocol
adopting $K$-means clustering, when varying the sample size. These percentages are evaluated establishing a sig-
nificance level of 5% for the compatibility test. The number of cluster was 10.
References


Quantum Hamiltonian learning using Bayesian inference on a quantum photonic simulator

J. Wang1, S. Paesani1, R. Santagati1, S. Knauer1, A. A. Gentile1, N. Wiebe2, M. Petruzzella3, A. Laing1, J. G. Rarity1, J. L. O’Brien1, and M. G. Thompson1

1 Quantum Engineering Technology Labs, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, BS8 1FD, UK.
2 Quantum Architectures and Computation Group, Microsoft Research, Redmond, Washington 98052, USA.
3 Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, NL-5600MB Eindhoven, The Netherlands.

Correspondence to: aa.gentile@bristol.ac.uk

Abstract: Using classical machine learning to control a silicon-photonics quantum simulator, we implemented the recently proposed quantum Hamiltonian learning (QHL) protocol, successfully emulating the electron spin dynamics in a NV-centre. We experimentally verified also an interactive version of QHL, using it to self-characterise our photonic quantum simulator. These results show the feasibility of efficient characterisation and verification of quantum devices and systems.

1. Introduction

The behaviour of a physical system can often be described in terms of a model Hamiltonian. Learning it is important for the understanding, simulating and characterising the system. In the quantum realm this is key for a vast variety of studies, ranging from foundational to applicative ones [1, 2]. However, classical computers are expected to be inadequate to solve this problem for exponentially large instances [3], leaving open the question of how to efficiently characterise and verify complex quantum systems. The advent of quantum simulators [4] recently inspired the proposal of a new method: quantum Hamiltonian learning [5]. QHL enhances the simulator capabilities via classical machine learning, enabling the efficient learning and validation of Hamiltonian models also for quantum systems. The exponential speed-up in the inference process is triggered in two ways. First, the adoption of a quantum device to simulate and eventually reproduce the dynamics of another quantum system prevents the curse of dimensionality that would affect a classical simulator. This is at the heart of the Quantum Likelihood Estimation protocol (QLE). In addition, whenever a quantum channel between the system and the simulator can be leveraged upon, an Interactive Quantum Likelihood Estimation (IQLE) version of the protocol can be invoked, further enhancing the speed-up in the learning process [5].

Here we experimentally demonstrate QLE using a programmable silicon-photonics quantum simulator (Fig.1a) to estimate the best Hamiltonian parameters (among those accessible by the simulator), in order to reproduce the electron spin dynamics of a negatively charged nitrogen-vacancy (NV−) centre in bulk diamond (Fig.1b).

2. Results

We adopt as quantum simulator a controllable Si quantum photonic chip [6, 7]. (Fig.1a) The signal and idler photons are generated by spontaneous four-wave mixing and used to prepare the path-encoded maximally entangled state [6]. The idler photon passes through a state preparation $|\phi\rangle$ stage and an unitary, $\hat{U}$ or $\hat{V}$, controlled by the state of the signal [7]. The quantum operations are achieved by thermo-optical phase gates. Projective measurements $\hat{M}$ allow us to estimate the likelihoods function for the QHL implementations.

We study the dynamics of the negatively charged NV− centre’s ground-state electron spin [8, 9], see Fig.1b, focusing on the transition between the states $m_s = 0$ and $m_s = -1$ of the ground-state triplet (Fig.1d). The electron spin state is addressed and read-out via a confocal microscope setup. The electron spin is optically initialised into the $m_s = 0$ state, and then driven in a single Rabi sequence (Fig.1c) by applying microwave pulses. The state occupancy probabilities at the end of the evolution are measured via the NV− centre photo-luminescence (PL). These are then fed into the quantum simulator to perform the estimation of the likelihood function $Pr(D|\vec{x})$ from the measurement statistics obtained when
Fig. 1. a. Schematic of a silicon quantum photonic simulator. b. Schematic of a diamond nitrogen-vacancy centre. c. Initialisation, manipulation and read-out of the electron spin state. d. Energy-level diagram of a NV centre in diamond. e. Self-characterisation of the photonic chip via IQLE. The quadratic loss is referred to the true parameter $f_0$. f. The learning of the spin Hamiltonian represented by a rescaled Rabi frequency $\omega$.

dialling in the simulator a Hamiltonian $\hat{H}(f)$. The spin Hamiltonian studied here can be modelled as $\hat{H}(f) = \hat{\sigma}_x f / 2$. Iterating the Bayesian update process [5] allows to infer the Hamiltonian parameter, that is here the Rabi frequency of the NV spin. Fig. 1f shows the learning of a rescaled Rabi frequency $\omega$, progressively approaching the correct value of $\omega_0$. The final value learnt by QLE is $f = (6.93 \pm 0.09)\text{MHz}$, in good agreement with the one obtained from a curve fit of Rabi oscillations.

In our IQLE implementation, the unitary $\hat{U} = \exp(-i f_0 \hat{\sigma}_x / 2)$ implemented in the photonic chip acts as a non-characterised part of the quantum device, connected to the trusted simulator implemented as $(\hat{V})$ via an intrinsically quantum channel. For this configuration, we show in Fig. 1e that IQLE can efficiently characterise $\hat{U}$, by learning $f_0$ exponentially fast.

3. Conclusions

We have reported experimental implementation of quantum Hamiltonian learning protocols, using (imperfect) photonic quantum simulators and NV$^-$ centres in bulk diamond. This work shows a new approach for efficiently verifying real quantum physical systems and future large-scale quantum devices by enhancing established classical computational techniques with quantum processing power.

References

Supervised learning of time-independent Hamiltonians for gate design

Luca Innocenti,1 Leonardo Banchi,2 Alessandro Ferraro,1 Sougato Bose,2 and Mauro Paternostro1,3

1Centre for Theoretical Atomic, Molecular, and Optical Physics, School of Mathematics and Physics, Queen’s University Belfast, BT7 1NN Belfast, United Kingdom
2Department of Physics and Astronomy, University College London, Gower Street, WC1E 6BT London, United Kingdom
3Laboratoire Kastler Brossel, ENS-PSL Research University, 24 rue Lhomond, F-75005 Paris, France

Short abstract: We present a general framework to approach the problem of finding time-independent dynamics generating target unitary evolutions. More specifically, given a target unitary gate $\mathcal{G}$ over a set of qubits, and a parametrized Hamiltonian of the form $\mathcal{H}(\lambda) = \sum_i \lambda_i \sigma_i$ with $\sigma_i$ Hermitian operators over the qubits, we want to find a set of values $\lambda_0$ such that $\exp(i\mathcal{H}(\lambda_0)) = \mathcal{G}$. We show that this problem is equivalently stated as a set of conditions over the spectrum of $\mathcal{H}(\lambda)$, reducing the problem to an inverse eigenvalue problem. We show how to solve this problem in some physically relevant instances, like for example to find time-independent dynamics implementing Toffoli and Fredkin gates without the need for ancillary gates or effective evolutions. More generally, we show how our framework can be used as a standard point for a numerical supervised learning technique to solve the same problem. Using the latter, we are able to find solutions for other cases which are harder to solve analytically. For example, we can use this technique to find generators for the Toffoli gate using only diagonal pairwise interactions, easier to implement in some experimental architectures.

Extended abstract: Among the fundamental requirements for most quantum algorithms is the capability of implementing a given experimental platform non-trivial entangling gates between sets of qubits, like for example Toffoli (controlled-controlled-not) or Fredkin (controlled-swap) gates. Doing this is however often non-trivial, as the dynamics generating these kinds of gates requires types of interactions, like three-qubit interactions, that may not be naturally implementable in a given experimental architecture. A common way around this problem is to decompose the “hard” gates into series of simpler ones that can be directly implemented. For example, decomposing a Toffoli gate using five two-qubit gates [1]. These kind of decompositions are however not ideal, as the fidelity of the target gate decreases with the number of intermediate gates that must be used. Another possibility is to implement the gates using a quantum control approach, that is, using time-dependent dynamics. Some recent works showcased this approach for gates like Toffoli and Fredkin [2-4].

Using a different approach, Banchi et al. [5] recently showed that using an ancillary gate, it is possible to implement these gates using time-independent dynamics, that is, by simply fixing a specific set of pairwise interactions among four qubits and waiting a fixed amount of time, tracing out the ancillary qubit at the end. The parameters generating this kind of evolution were found using a supervised learning approach similar to the ones commonly used for some classes of machine learning algorithms. Building on this work, we develop a general framework to more easily solve the problem of finding a time-independent dynamics resulting in a given target gate. This framework enables us to look for analytical solutions in some simple but physically relevant instances, but can also be used more generally as a starting point for the numerical optimization via supervised machine learning.

In this work we showcase both applications of our framework. Firstly, we provide exact, explicit Hamiltonians that generate Toffoli and Fredkin gates using only pairwise interactions, thus proving that it is not necessary to use time-dependent dynamics nor ancillary qubits to obtain such gates. This is an interesting fact on its own, as one could have naively though this to be not possible even in principle. For instance, we find the following Hamiltonian $\mathcal{H}_{\text{Fred}}$, containing only diagonal pairwise (and single-qubit) interactions:

$$\mathcal{H}_{\text{Fred}} = \frac{\pi}{8} \sqrt{\frac{143}{5}} (\sigma_2^x + \sigma_3^z) + \frac{5\sqrt{3}\pi}{8} (\sigma_1^x \sigma_2^x + \sigma_1^z \sigma_3^z)$$
$$+ \frac{3\pi}{4} \sqrt{\frac{7}{5}} (\sigma_1^x \sigma_2^z + \sigma_1^z \sigma_3^z) + \frac{\pi}{2} \sigma_1^z$$
$$- \frac{3\pi}{8} (\sigma_2^x \sigma_3^z + \sigma_2^z \sigma_3^x + \sigma_2^x \sigma_3^z) + \frac{3\pi}{8} \mathbb{I}.$$  \hspace{1cm} (1)

As can be verified directly, the above satisfies $\exp(i\mathcal{H}_{\text{Fred}}) = \mathcal{U}_{\text{Fred}}$, and proves that diagonal pairwise interactions are enough to implement a Fredkin gate without decomposition or time-dependent control. Similarly, we find a class of generators for the Toffoli gate. One of these generators is for example the following:

$$\mathcal{H}_{\text{Toff}} = \frac{\pi}{8} \left[ 5 - 2\sigma_3^z - \sigma_1^z - \sigma_2^z + (\sigma_1^z + \sigma_2^z)\sigma_3^x - 3\sigma_1^z \sigma_2^z + \sqrt{15} (\sigma_2^z - \sigma_1^z) \sigma_3^z \right].$$  \hspace{1cm} (2)
which is easily verified to satisfy $\exp(iH_{\text{Toff}}) = U_{\text{Toff}}$, but contains non diagonal interactions of the form $\sigma^z_i$.

On the other hand, when solutions cannot be found analytically, we rely on a numerical approach inspired by supervised learning techniques. We can reframe the above described gate design problem as an optimization problem as following: given a target gate $G$, find the set of real parameters $\lambda$ such that $F(\lambda, \psi) \equiv \langle \psi | \exp(iH(\lambda)) | \psi \rangle$ is maximum, for all states $|\psi\rangle$, or, equivalently, find the set of real parameters $\lambda$ such that $F(\lambda, \psi) = 1$ for all $|\psi\rangle$. The direct approach of using standard optimization techniques to maximize this fidelity averaged over the states, $\bar{F}(\lambda)$, reveals to be inefficient due to the complexity of the parameter space. A supervised learning approach enables us instead to use the variability of the parameter landscape with respect to the states $|\psi\rangle$ as an advantage, rather than as a hindrance. In particular, we implement a gradient-descent based technique, which can be broken down into the following 5 steps:

1. Choose an initial set of parameters $\lambda$ (randomly or using some physically-inspired Ansatz).

2. Generate a random set of input states $|\psi_k\rangle$, $k = 1, ..., N_b$, with $N_b$ the size of the mini-batches chosen beforehand. The value of the learning rate is furthermore chosen to be decreasing with the iteration number.

3. For each $k$, compute $\nabla_\lambda F(\lambda, \psi_k)$.

4. Update the coupling strengths $\lambda$. We here do this using the so-called momentum gradient descent method [6], corresponding to the following updating rule:

$$
\begin{align*}
\eta & \rightarrow \gamma \eta + \eta \nabla_\lambda F(\lambda, \psi_k), \\
\lambda & \rightarrow \lambda + \eta \nabla_\lambda F(\lambda, \psi_k),
\end{align*}
$$

where the learning rate $\eta$ and the momentum term $\gamma$ are hyperparameters to be chosen beforehand. Alternative methods, like ADAGRAD or ADADELTA [6], can be used and can given better or worse performances, depending on the circumstances.

5. Return to point 2, until a satisfactory value of the fidelity is obtained.

A possible bottleneck in this algorithm is the third step above: the computation of the gradient of the fidelity with respect to the parameters. This was indeed one of the limiting factors, performance-wise, of the previous implementation of this approach [5]. The problem arises because directly calculating the derivative, using for example a finite difference approximation, is computationally expensive and does not scale well with the number of parameters to be maximized. We here overcome this problem using an Automatic Differentiation algorithm [7, 8]. This is a technique often employed in machine learning techniques that allows to, loosely speaking, compute the derivative of functions expressed as algorithms. To do this we implement the cost function of our problem (in our case the fidelity function shown above) as a computational graph, that is, as a nested sequence of elementary operations the gradient of which is known analytically. Doing this allows to drastically improve the scaling of the whole optimization procedure, and explore a much broaded set of scenarios. Using this and other optimization techniques we are able to optimize qubit networks with 8 or more qubits, over hundreds of interaction parameters. This approach is also highly flexible, being rather transparent to the particular form of the used cost function. This means that while we here mainly showcased how this algorithm can be used to find parameters implementing specific unitary evolutions, it can also be used to optimize networks with ancillae, or to optimize networks to implement specific open dynamics, rather than only unitary ones.

Challenges to Practical End-to-end Implementation of Quantum Optimization Approaches for Combinatorial problems (Invited Talk)

Davide Venturelli
USRA/NASA
Quantum Artificial Intelligence Laboratory

We discuss the hurdles encountered for programming and compiling solvers of real world combinatorial optimization problems. We will compare the overhead and resource requirements of running problems such as coloring, scheduling, knapsack into realistic quantum annealer architectures (embedding problem) with the one for compiling these problem instances on gate-model devices within the quantum alternate operator ansatz / quantum approximate optimization (QAOA) approach.
Annealing Computation with Adaptor Mechanism and its Application to Property-Verification of Neural Network Systems

National Institute of Advanced Industrial Science and Technology (AIST)

With the on-going infiltrations of artificial intelligence (AI) technology into our daily lives, developments of verification technology for its safety and reliability will be more significantly required than ever. Unlike conventional software/hardware, AI (or neural network (NN)) obtained by a learning process is often based on empirical data rather than humanly understandable logics. Because of the difference, verification technology for the conventional systems merely works restrictively when we apply it to the new technology. We need something more powerful for the verification of the new technology. Meanwhile, developments of quantum information technologies are remarkable today. In particular, quantum annealing computation is expected to be one of the first practical quantum computations that possibly brings further computation power than existing classical one. With the above circumstance, assuming existing of practical quantum annealing system, we theoretically discuss some applications of the quantum annealing computation to property-verification of given NN systems.

We show constructions of problem Hamiltonians \( \hat{H}_P \) in Ising Model (including two body interactions at most) that can efficiently decide True/False, by applying quantum annealing, of propositions such as:

Example:

Given a neural network \( y = N(x) \) and a particular output \( y^* \), there exists \( x^* \) satisfying \( N(x^*) = y^* \).

To understand an implication of the exampled proposition in a context of the verification, imagine NN, which was well-trained by some learning data, and is to be used in the automated driving system. Input \( x \) is a signal from the sensor, and operation action is decided by output \( y \). Suppose \( y^* \) corresponding to an inappropriate operation (like a simultaneous action on brakes and accelerator). When the proposition is verified to be false, the neural network is proven intrinsic safety in the sense that there is no input related to the inappropriate operation.

To achieve the construction of such problem Hamiltonian \( \hat{H}_P \), there are two underlying ideas: one is NN-Hamiltonian whose degenerated ground-states give parallel simulation of all possible behavior of the given NN-system, another is introduction of the Adaptor Mechanism. In the following, we explain the ideas along the above exampled proposition.

**NN-Hamiltonian:** For any neural network \( y = N(x) \) which is describable in combinational logic, we can always find a Ising-Model Hamiltonian [1,2] of form of

\[
\hat{H}_{NN} := \sum_j h_j \sigma_j^{(x)} + \sum_{j<k} J_{jk} \sigma_j^{(x)} \sigma_k^{(x)}
\]

whose degenerated ground-states are

\[\{ |x\rangle \otimes |N(x)\rangle \otimes |A(x)\rangle \}_{x \in X} \in \mathcal{H}_{NN}\]

where

\[\mathcal{H}_{NN} := \mathcal{H}_x \otimes \mathcal{H}_y \otimes \mathcal{H}_{ancilla} \cong \mathbb{C}^{\log_2|\chi|} \otimes \mathbb{C}^{\log_2|\chi|} \otimes \mathbb{C}^{|\chi|} \otimes \mathbb{C}^{|\chi|}\]

\(\chi \) and \(\chi'\) are the domain and the range of \(N(x)\) respectively. (The integer \(d\) is the number of ancilla qubits required to construct such \(\mathcal{H}_{NN}\), and is enough to be polynomial with respect to \(\log_2|\chi|\). In this sense, we can always find an efficient \(\hat{H}_{NN}\).) When quantum annealing is applied to \(\mathcal{H}_{NN}\), the final state to be obtained is a superposition state of all of the degenerated ground-states. In other words, in the superposition state, all possible behavior of the neural network is “parallelly simulated”. In addition, let us introduce \(\hat{H}_y\) whose ground-state is \(|y\rangle \in \mathcal{H}_y \cong \mathbb{C}^{\log_2|\chi'|}\). Using a binary representation of \(y^* := (y_1^*, \ldots, y_{|\chi'|}^*)\) with \(y_k^* \in \{0,1\}\), \(\hat{H}_y\) can be immediately given as

\[\hat{H}_y := -\lambda \sum_{k=1}^{\log_2|\chi'|} y_k^* \sigma_k^{(x)} \text{ with } \lambda > 0 \text{ and Pauli matrices } \sigma_k^{(x)} \text{ on } \mathcal{H}_y.\]

Similarly, let us introduce a ferromagnetic interaction between corresponding qubits on \(\mathcal{H}_x \otimes \mathcal{H}_y\) as

\[\hat{H}_F := \]
\[-J \sum_{k=1}^{ \log_2 |x|^{(z)}} \sigma_k^{(z)} \otimes \sigma_k^{(z)} \] with \( J > 0 \). Ground state obtained by annealing process with respect to a problem Hamiltonian

\[ H_{NN}^{F} := \hat{H}_{NN} \otimes \hat{I}_{\hat{H}_x} + \hat{I}_{\hat{H}_{\text{ancilla}}} \otimes \hat{H}_F \] (4)

gives \( x^* \) that holds \( \mathcal{N}(x^*) = y^* \) unless

\[ D := \dim \left( \mathcal{G} (\hat{H}_{NN} \otimes \hat{I}_{\hat{H}_x}) \cap \mathcal{G} (\hat{I}_{\hat{H}_{\text{ancilla}}} \otimes \hat{I}_{\hat{H}_y}) \right) = 0 \] (5)

where \( \mathcal{G}(\hat{H}) \) is a Hilbert space spanned by the ground state(s) of \( \hat{H} \). If eq.(5) holds, it means that there does not exist \( x^* \) such as \( \mathcal{N}(x^*) = y^* \).

**Adaptor Mechanism:** We introduce additional degree of freedom (or single qubit on Hilbert space \( \mathcal{H}_A \)) to the annealing process to make it possible to physically observe whether eq.(5) holds or not. For simplicity, let us consider the case where \( y \) is one-bit variable. (Extension to multi-bit cases can be straightforwardly obtained.) Just like an adaptor, we insert the new qubit between two qubits on \( \mathcal{H}_y \otimes \mathcal{H}_y \) by introducing new interactions

\[ \hat{H}_{AM} := -J \sigma^{(z)} \otimes \sigma^{(z)} - \gamma \hat{I}_{\hat{H}_x} \otimes \sigma^{(z)} \otimes \sigma^{(z)} \] on \( \mathcal{H}_y \otimes \mathcal{H}_A \otimes \mathcal{H}_y \) (6)

instead of \( \hat{H}_F \) on \( \mathcal{H}_y \otimes \mathcal{H}_y \). The positive constant \( \gamma \) is chosen to be

\[ 0 < \gamma < \Delta E \] (7)

where \( \Delta E \) is a gap between the first excited level and the ground level of Hamiltonian in eq.(4). From the construction of the Hamiltonian in eq. (4), estimation of \( \Delta E \) can be immediately obtained. In the case of \( D \neq 0 \), concerning a problem Hamiltonian

\[ \hat{H}_P = \hat{H}_{NN} \otimes \hat{I}_{\hat{H}_x} + \hat{I}_{\hat{H}_{\text{ancilla}}} \otimes \hat{H}_y \] (8)

\( \mathcal{G}(\hat{H}_P) \) is a Hilbert space spanned by

\[ \{ |x^* \rangle \otimes |y^* \rangle \otimes |A(x^*) \rangle \rangle \} \text{ s.t. } \mathcal{N}(x^*_y \rangle = y^* \in \mathcal{H}_x \otimes \mathcal{H}_y \otimes \mathcal{H}_{\text{ancilla}} \otimes \mathcal{H}_A \otimes \mathcal{H}_y. \] (9)

Otherwise, in the case of \( D = 0 \), \( \mathcal{G}(\hat{H}_P) \) is a Hilbert space spanned by

\[ \{ |x^{**} \rangle \otimes |y^{**} \rangle \otimes A(x^{**}) \rangle \rangle \} \text{ s.t. } \mathcal{N}(x^{**}_y \rangle = y^* \rangle \oplus 1 \in \mathcal{H}_x \otimes \mathcal{H}_y \otimes \mathcal{H}_{\text{ancilla}} \otimes \mathcal{H}_A \otimes \mathcal{H}_y. \] (10)

In the latter case, one can find that the ferromagnetic relation between the two qubits on \( \mathcal{H}_A \otimes \mathcal{H}_y \) is broken to maintain the simulated behavior of the neural network in the ground states. (In order to obtain this mechanism, the condition on \( \gamma \) in eq.(7) is required.) Equations (9) and (10) show that, by applying quantum annealing to \( \hat{H}_P \) in eq.(8) and by observing the state of the qubit on \( \mathcal{H}_A \), one can decide **True/FALSE** of the exampled proposition.

Besides the above, we can apply our approach to decide True/FALSE of propositions such as

**Example2:** Given two neural networks \( \mathcal{N}(x) \) and \( \mathcal{N}^*(x) \), \( \mathcal{N}^*(x) = y^* \) holds for all \( x \) satisfying \( \mathcal{N}(x) = y^* \), showing an inclusion relation among the two neural networks, which is essential in verifying a property of one neural network in comparison with another one [3]. We will present some more details on the approach and an example of the concrete form of \( H_{NN} \) as well.


Robust quantum optimizer with full connectivity

Simon E. Nigg, Niels Lörch, and Rakesh P. Tiwari
Department of Physics, University of Basel, Switzerland

Quantum phenomena have the potential to speed up the solution of hard optimization problems. For example, quantum annealing, based on the quantum tunneling effect, has recently been shown to scale exponentially better with system size than classical simulated annealing. However, current realizations of quantum annealers with superconducting qubits face two major challenges. First, the connectivity between the qubits is limited, excluding many optimization problems from a direct implementation. Second, decoherence degrades the success probability of the optimization. We address both of these shortcomings and propose an architecture in which the qubits are robustly encoded in continuous variable degrees of freedom. By leveraging the phenomenon of flux quantization, all-to-all connectivity with sufficient tunability to implement many relevant optimization problems is obtained without overhead. Furthermore, we demonstrate the robustness of this architecture by simulating the optimal solution of a small instance of the nondeterministic polynomial-time hard (NP-hard) and fully connected number partitioning problem in the presence of dissipation.
Adiabatic quantum optimization applied to the stereo matching problem

William Cruz-Santos
Computer Engineering, CU-UAEM Valle de Chalco
Estado de México, México

Salvador E. Venegas-Andraca
Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias
Estado de México, México

Marco Lanzagorta
US Naval Research Laboratory, Washington, USA

Quantum phenomena have the potential to speed up the solution of hard optimization problems. For example, quantum annealing, based on the quantum tunneling effect, has recently been shown to scale exponentially better with system size than classical simulated annealing. However, current realizations of quantum annealers with superconducting qubits face two major challenges. First, the connectivity between the qubits is limited, excluding many optimization problems from a direct implementation. Second, decoherence degrades the success probability of the optimization. We address both of these shortcomings and propose an architecture in which the qubits are robustly encoded in continuous variable degrees of freedom. By leveraging the phenomenon of flux quantization, all-to-all connectivity with sufficient tunability to implement many relevant optimization problems is obtained without overhead. Furthermore, we demonstrate the robustness of this architecture by simulating the optimal solution of a small instance of the nondeterministic polynomial-time hard (NP-hard) and fully connected number partitioning problem in the presence of dissipation.
Opportunities and Challenges for Quantum-Assisted Machine Learning in Near-Term Quantum Computers

Alejandro Perdomo Ortiz
Quantum Artificial Intelligence Laboratory
NASA Ames Research Center

Abstract

With quantum computing technologies nearing the era of commercialization and quantum supremacy, machine learning (ML) appears as one of the promising “killer” applications. Despite significant effort, there has been a disconnect between most quantum machine learning proposals, the needs of ML practitioners, and the capabilities of near-term quantum devices to demonstrate quantum enhancement in the near future. In this talk, we provide concrete examples of intractable ML tasks that could be enhanced with near-term devices. We argue that to reach this target, the focus should be on areas where ML researchers are still struggling, such as generative models in unsupervised or semisupervised learning, instead of the popular and much more tractable ML techniques. We also highlight the case of classical datasets with potential quantum-like statistical correlations where quantum models could be more suitable. We focus on hybrid quantum-classical approaches and illustrate some of the key challenges we foresee for near-term implementations. As a concrete framework, we introduce the quantum-assisted Helmholtz machine (QAHM); an attempt to use near-term quantum devices to tackle high-resolution datasets on continuous variables. Instead of using quantum computers to assist deep learning, as previous approaches do, the QAHM uses deep learning to extract a low-dimensional binary representation of data, suitable for relatively small quantum processors which can assist the training of an unsupervised generative model. To demonstrate this concept on a real-world dataset, we used 1644 quantum bits of a noisy non-fault-tolerant quantum device, the D-Wave 2000Q, to assist the training of a sub-sampled version of the MNIST handwritten digit dataset with 16 x 16 continuous valued pixels. Although we illustrate this concept on a quantum annealer, adaptations to other quantum platforms, such as ion-trap technologies or superconducting gate-model architectures, could be explored within this flexible framework.
Distinguishing free and interacting as pattern recognition

Christopher J. Turner, Konstantinos Meichanetzidis, Zlatko Papi and Jiannis K. Pachos

University of Leeds

A physical theory is free if the behaviour of a collection of many particles can be constructed from the behaviour of a single particle in isolation. Free theories are important for our understanding of the general interacting case not only as a source of intuition. They appear as the renormalised theory of quasi-particles around which a perturbative expansion is found even in strongly-correlated systems. In some cases these quasi-particles exotic exhibit anyonic statistics at low energies – a hallmark of emergent topological order.

The pattern of many-body quantum entanglement in a free system has a description in terms of a reduced number of degrees of freedom in the two-point correlations. The ‘interaction distance’ [Nat. Commun. 8, 14926 (2017)] exploits this fact to classify states according to their distinguishability from the class of free states by the information contained in its entanglement. Associated with the interaction distance is the particular free state which is optimal among this class for approximating the input state which we term the optimal free model.

We provide an introduction to the theory of the interaction distance, discussing its universal scaling behaviour close to criticality and in gapped phases its convergence to a limit characteristic of the phase. At its heart calculating \( D_F \) and finding the optimal free model involves the solution of a non-convex global optimisation problem. We give our previous strategies for selecting an initial guess from which we use a Monte-Carlo basin hopping strategy to ensure that we find the global minimum. Here we apply supervised learning to produce an improved initial guess from which we can forgo the expense of a global optimisation strategy such as Monte-Carlo.

We will also present preliminary work on using supervised learning to approximate the ground state interaction distance directly from the Hamiltonian, circumventing the computationally difficult step of first finding the ground state. We note that it would also be interesting to use machine learning to find the optimal free model directly from the Hamiltonian, transforming the interaction distance from a diagnostic tool into a practical approximation method.
Free-fermion descriptions of parafermion chains

Konstantinos Meichanetzidis, Christopher J. Turner, Ashk Farjami, Zlatko Papic and Jiannis Pachos

University of Leeds

Majorana fermions have been of major interest in the last decade both due to their fundamental interest and their relevance to quantum technologies. They arise as quasiparticles at the edges of one-dimensional superconducting wires as experimental results strongly support. Majoranas exhibit non-Abelian statistics and so, when braided in a wire-network, they can perform unitary gates, thus implementing quantum computation, though non-universal. Their efficient manipulation in the lab is also under-way. The simplest model in 1D hosting Majoranas is the archetypal example of the Kitaev wire, which in many cases is analytically tractable. It is described by a free-fermion Hamiltonian and thus it is efficiently simulatable.

Parafermions are defined as generalisations of Majoranas. They are emergent quasiparticles in parafermion chains and experimental proposals are promising for realising a Fibonacci phase by coupling them, which in principle suffices for universal quantum computation by braiding. Majoranas are viewed as the simplest parafermions of order 2, and appear as zero-modes at the edges of the chains. In constrast with the Majorana case, when the order of the parafermions is increased, the chain cannot be described by free-fermions and are thus intrinsically interacting in terms of fermions. Then a natural question arises; how important are interactions for the implementation of parafermions? It is then in our interest to develop reliable diagnostics of the effect that interactions have on the ground state of a Hamiltonian.

We therefore to the ‘interaction distance’ [Nat. Commun. 8, 14926 (2017)], a measure of the Gaussianity of the correlations present in a state. The interaction distance takes as input the entanglement spectrum after a state is bi-partitioned and optimises for the closest Gaussian entanglement spectrum according to the trace distance. In this way, a free-fermion description of the state is determined. Computing the interaction distance for the parafermion chains leads to surprising results about their ground states. Interestingly, analytic results are possible at the renormalisation fixed point for all parafermion orders. In particular, we find that chains of order equal to some power of 2, though described by intrinsically interacting Hamiltonians, have ground states that show free-fermion structure in their correlations. Furthermore, it is possible to find a local parent Hamiltonian, whose energy spectrum differs from that of the parafermion chain’s, but generates the same ground state. Exploration
of low lying excited states showed that the interaction distance of chains whose order matches that of coupled Majorana chains also vanishes. In addition, we find that chains incompatible with coupled Majorana wires achieve a finite interaction distance, with some of them almost saturating the maximal possible value. These results establish the interaction distance as a novel diagnostic of interactions in these exotic systems.

Determining the interaction distance entails a non-convex optimisation problem. The algorithm performs basin-hopping together with a common local minimisation procedure in order to approximate the global minimum over all Gaussian spectra. As it stands, even if the optimisation parameters grow only linearly with system size, both the input entanglement spectrum and the search space scale exponentially in general. Neural networks under supervised and unsupervised learning, or re-enforcement learning have been established as the main tool to tackle pattern matching and pattern recognition tasks efficiently. Here, we present preliminary results for discriminating Gaussian correlations (or free-fermion-like) from non-Gaussian ones using machine learning on ground states of parafermion chains. Machine learning techniques can be used in order to identify patterns of correlations of effective quasiparticles emerging from the strongly correlated parafermion systems that were not expected before, as is the case of the results obtained by application of the interaction distance. The ground states themselves can be efficiently obtained via matrix-product-state methods. However, as the problem of finding the ground state is another optimisation problem, we use re-enforcement learning in order to obtain the ground state. In principle, a neural network could be constructed in order to obtain directly the interaction distance from the Hamiltonian by performing both tasks in an integrated way.
Identifying Free Particle Correlations in Topologically Ordered Systems

Ashk Farjami
University of Leeds

Topologically ordered systems display many interesting ground state properties and excitation statistics - fractionalisation, ground state degeneracies and exotic excitations that posses anyonic statistics. These properties, typically arising in the strongly correlated fractional quantum Hall liquids, are thought to be linked to interactions between particles. Interaction distance, [Nat. Commun. 8, 14926 (2017)], is a measure of the interactiveness of a state in terms of its distance from a Gaussian state of free fermions. Free fermion systems have certain patterns in their many-body entanglement spectra. We use interaction distance as a pattern recognition tool to detect the similarities between these entanglement spectra and those of certain presumed interacting topological phases of matter. This measure allows us to assess the necessity of interactions in the emergence of topological features, such as the ground state degeneracies and exotic fusion and braiding statistics that make these models such an exciting focus of study.

2D-String-nets and Walker-Wang models (a 3D generalisation) are quantum systems of extended objects that give rise to these topological phases of matter. Using interaction distance we study the entanglement spectra of all Abelian $\mathbb{Z}_N$ string-nets and Walker-Wang models. We find that the quantum correlations of a certain class, specifically $\mathbb{Z}_{2^n}$ models, can be described exactly by the quantum correlations of free fermions in both the 2D and 3D cases. All other $\mathbb{Z}_{N\neq2^n}$ models saturate the maximum possible interaction distance from their optimal free-fermion description. We also study non-Abelian models ($SU(2)_k$) and find none with quantum correlations describable by free fermion quantum correlations.

Here, we fully detail the transformation between the class of $\mathbb{Z}_{2^n}$ gauge theory to spin models and then through a generalised Jordan-Wigner transformation to their free fermion descriptions. We show how interaction distance has allowed us to establish new types of fermionisation procedures to describe the physics of 2D and 3D topologically ordered systems and helped us take the next step in the pursuit of experimental realisations of these models.

The development of these fermionisation procedures becomes more arduous as we increase the complexity of the model we are studying (i.e. by increasing the dimension of the symmetry group of the theory, or the spacial dimension...
we are working in). We are interested in the potential machine learning has to produce efficient neural networks able to find transformations from gauge theories to their free fermion descriptions.

We would also like to use quantum machine learning pattern recognition techniques, to extend this study to recognising patterns not only of free fermions but of other exotic quasiparticles. This would be with the aim of decomposing complex topological theories in terms of simpler topological theories, as we have done transforming the $Z_{2^n}$ models to normal free fermions.
Recognizing complex patterns is a central problem which pervades all fields of science. The increased computational power of modern computers has allowed the application of advanced methods to the extraction of such patterns from humongous amount of data and we are witnessing an ever increasing effort to find novel applications in numerous disciplines. Quantum physicists seized the opportunity and started to develop new ideas in the direction of what is now called Quantum Machine Learning\(^1\). This line of research is divided in two main different branches. The first tries to develop quantum algorithms capable of learning, i.e. to exploit speed ups from quantum computers to make machines learn faster and better. The second, that we will consider in this work, tries to use classical machine learning algorithms to extract insightful information about quantum systems.

The versatility of machine learning has allowed scientists to employ it in a number of problems which span from quantum control\(^2\)–\(^4\), where for instance one aims to design the optimal quantum circuit for achieving a given transformation, to tomography\(^5\), where the most likely quantum state has to be constructed from an incomplete set of data. In the last few years we are experiencing interesting developments also for some central problems in condensed matter, such as quantum phase classification/recognition\(^6\)–\(^9\), improvement of dynamical mean field theory\(^10\), enhancement of Quantum Monte Carlo methods\(^11\),\(^12\) or approximations of thermodynamical observables in statistical systems\(^13\).

An idea which received a lot of attention from the scientific community consists in using neural networks as variational wave functions to approximate ground states of many-body quantum systems\(^14\)–\(^16\). These networks are trained/optimized by the standard Variational Monte Carlo (VMC) method and while a few different neural networks architectures have been tested\(^14\)–\(^16\), the most promising results so far have been achieved with Boltzmann Machines\(^17\). In particular, state of the art numerical results have been obtained on popular models with Restricted Boltzmann Machines (RBM) and recent effort has demonstrated the power of Deep Boltzmann Machines to represent ground states of many-body Hamiltonians with polynomial-size gap and quantum states generated by any polynomial size quantum circuits\(^18\),\(^19\).

Other seemingly unrelated classes of states that are widely used in condensed matter physics are Tensor Networks States. In 1D, Matrix Product States (MPS) can approximate ground states of physical Hamiltonians efficiently\(^20\),\(^21\), and their structure has led to both analytical insights over the entanglement properties of physical systems as well as efficient variational algorithms for approximating them\(^22\)–\(^23\). The natural extension of MPS to larger dimensional systems are Projected Entangled Pair States (PEPS)\(^24\), which have been successful in describing strongly correlated quantum many body systems at zero and finite temperature. Their exact contraction is however \#P hard\(^25\) and algorithms for optimizing them need to rely on approximations.

Another approach to define higher dimensional Tensor Networks consists in first dividing the lattice into overlapping clusters of spins. The wave function of the spins in each cluster is then described by a simple Tensor Network. The global wave function is finally taken to be the product of these Tensor Networks, which introduces correlations among the different clusters. This construction for local clusters parametrized by a full tensor gives rise to Entangled Plaquette States (EPS)\(^26\)–\(^28\), while taking one dimensional clusters of spins each described by a MPS leads to a String-Bond States (SBS) Ansatz\(^29\),\(^30\). These states can be variationally optimized using the VMC method\(^30\)–\(^31\) and have been applied to 2D and 3D systems.

It has been shown that local tensor networks can be represented by Deep Boltzmann Machines\(^18\),\(^19\), but in the general case the wave function itself of a Deep Boltzmann Machine is intractable, which means that different optimization algorithms have to be used. In this work we show that there is a stronger relation between Restricted Boltzmann Machines and Tensor Network States that can be optimized using the VMC method: short-range RBM, which include analytical examples such as the toric code\(^32\), any graph state, cluster states and coherent thermal states\(^18\), are a special subclass of EPS. While EPS are more general than short-range RBM and more stable to optimize, the size of local cluster they can include is more limited. We therefore argue that EPS or fully connected RBM should be preferred to short-range RBM for numerical purposes. Moreover, we show that fully connected RBM are a subclass of String-Bond States with a particularly flexible non-local geometry and low bond dimension. This gives additional insights over the geometric structure of RBM and provides a way of generalizing them to non-local String-Bond States with larger bond dimension and fixed geometry. This generalization also allows for the description of states with larger local Hilbert space. We give examples where SBS provide a more efficient representation of a ground state wave function than RBM, as well as examples where the opposite is true, thus showing that each class of states has distinct advantages and drawbacks. The flexibility of SBS allows for combining RBM and SBS in a common Ansatz while keeping the
advantages of each class of states and incorporating additional geometric information about the problem studied.

We then turn to a practical application on a challenging problem for traditional tensor networks methods, namely the approximation of a state with chiral topological order. Chiral topological states of matter break time-reversal symmetry and were first realized in the context of the Fractional Quantum Hall (FQH) effect\textsuperscript{34}, in which the electrons of a two-dimensional electron gas subject to a strong magnetic field form an incompressible quantum liquid supporting fractionally charged quasi-particle excitations. Significant progress has since been made towards the construction of lattice models displaying the same physics, either in Hamiltonians realizing fractional Chern insulators\textsuperscript{34–39} or in quantum anti-ferromagnets on several lattices\textsuperscript{40–42}. One approach to describe the wave function of these anti-ferromagnets is to use parton constructed wave functions\textsuperscript{43–46}. It has also been suggested to construct chiral lattice wave functions from the FQH continuum wave functions, the paradigmatic example being the Kalmeyer-Laughlin wave function, a lattice analog of the Laughlin wave function\textsuperscript{47}. One may wish for a more general class of Ansatz wave functions which can describe chiral topological states. Efforts to construct chiral topological states with PEPS have been undertaken recently\textsuperscript{48–51}. The resulting states are however critical and it has been proven in the non-interacting case that the local parent Hamiltonian of a chiral fermionic Gaussian PEPS has to be gapless\textsuperscript{49}.

In this work we show that such an obstruction does not carry on to Neural Networks Quantum States due to their non-local structure. Indeed, we prove that any Jastrow wave function, and thus the Kalmeyer-Laughlin wave function, can be written exactly as a RBM. While the resulting construction uses only polynomially many parameters, its scaling makes it impractical for numerical purposes. Nevertheless we show that a remarkable accuracy can be achieved with much less parameters than is required for an exact construction by optimizing a fully-connected RBM to approximate the ground state of a parent Hamiltonian of the Laughlin wave function. In addition, we numerically evaluate the power of EPS, SBS and RBM to approximate the ground state of a chiral spin liquid for which the Laughlin state is already a good approximation\textsuperscript{50}. We find that RBM and non-local SBS are able to achieve lower energy than the Laughlin wave function and that RBM are particularly efficient in the number of parameters they require compared to SBS. EPS and local SBS, on the other hand, have energies higher than the Laughlin state for the same number of parameters. We provide a method for combining RBM and SBS with an initial approximation of the ground state and apply this method by combining the Laughlin wave function with the previous Ansatz classes. This enables us to reach even lower energies and to characterize the properties of the ground state by computing its topological entanglement entropy, which is in agreement with a state sharing the topological properties of the Laughlin state.

\textsuperscript{3} M. August and X. Ni, Phys. Rev. A 95, 012335 (2017).
\textsuperscript{15} Z. Cai, arxiv:1704.05148 (2017).
\textsuperscript{19} Y. Huang and J. E. Moore, arxiv:1701.06246 (2017).
\textsuperscript{24} F. Verstraete and J. I. Cirac, arxiv:0407066 (2004).
\textsuperscript{27} F. Mezzacapo, N. Schuch, M. Bozincıgın, and J. I. Cirac, New J. Phys. 11, 083026 (2009).


Quantum Circuits for Quantum Channels

Raban Iten, Roger Colbeck, and Matthias Christandl
(Dated: 17th September 2017)

Introduction. Quantum channels, mathematically described by completely positive, trace-preserving maps, play an important role in quantum information theory because they are the most general evolutions quantum systems can undergo. In this work we consider how to best decompose them in terms of a simpler set of operations, specifically C-NOT and single-qubit gates, which are known to be universal [1]. By ‘best’ we broadly mean with the lowest experimental cost, which in this work we measure by the number of C-NOT gates.

Previous work has looked at the case of general unitary gates [2–7] and state preparation [6, 8], both of which are special cases of isometries. Recently, it was shown that every isometry from $m$ to $n$ qubits can be implemented using about twice the C-NOT count required by a lower bound [9]. Quantum operations beyond isometries have been investigated in [10], although without the goal of minimizing the number of C-NOTS, and in [11] the case of channels on a single qubit was considered. Here we consider C-NOT-efficient implementations of arbitrary quantum channels from $m$ to $n$ qubits.

Results. We give low-cost decomposition schemes for quantum channels from $m$ to $n$ qubits and lower bounds on the number of C-NOT gates in three different models. In the first, the Quantum Circuit Model (QCM), in addition to the single-qubit and C-NOT gates, qubits can be traced out. In the second (RandomQCM), we also allow external classical randomness. In the third (MeasuredQCM) we also allow measurements followed by operations that are classically controlled on the outcomes. We give near-optimal decompositions in almost all cases. Our results are summarized in Table I (where the first row of Table I was already known [9]). In all cases our decomposition schemes correspond to circuit topologies—fixed circuit structures with free parameters such that adjustment of these parameters enables any quantum channel from $m$ to $n$ qubits to be performed.

Our main result is a MeasuredQCM topology for any quantum channel from $m$ qubits to $n$ qubits that uses at most one ancilla and has a low C-NOT count. This decomposition generalizes [11] to arbitrary channels from $m$ to $n$ qubits. In fact, we recover a similar circuit topology (consisting of only one C-NOT gate) for single-qubit channels as given in [11] as a special case of our construction. Roughly, we decompose any quantum channel from $m$ to $n$ qubits into a short sequence of $m$ to $m + 1$ isometries (and an additional $m$ to $n$ isometry at the end of the sequence in the case $m < n$). Since this decomposition scheme uses the decomposition of isometries as a black box, it is not restricted to a specific universal gate library and hence we expect it to be useful in a wider range of contexts.

Applications. Experimental groups strive to demonstrate their ability to control a small number of qubits, and the ultimate demonstration would be to the ability to perform any quantum channel on them (cf. [12] and references therein). Using our decompositions, we derive low-cost

---

The technical version of this work is at arXiv:1609.08103 (and was published as Phys. Rev. A 95 052316 (2017)).

1 If $m = 0$ our channels allow the preparation of arbitrary mixed states.

2 The number of required $m$ to $m + 1$ isometries is at most $\lceil \log(K) \rceil$, where $K$ denotes the Kraus rank of the considered channel (and the logarithm is taken to base 2).
circuits for channels on a small number of qubits. In fact, every one to two qubit channel can be implemented by 4 C-not gates, every two to one channel by 7 C-not gates and every two to two channel by 13 C-not gates.

These counts are much lower than those known previously and are likely to be of relevance for experiments performed in the near future. For example, the best known implementation of a two to two channel in the QCM requires about 580 C-nots. If we also allow classical randomness, this count reduces to 54 C-nots, which is over four times the C-not count of 13 C-nots achieved by our decomposition scheme in the MeasuredQCM.

**Future work.** The long-term goal is to build a universal quantum compiler. In other words, to design an algorithm that takes as input a given set of gates, a noise model, an accuracy tolerance and a desired operation, and that gives as output a circuit composed of gates from the set that would approximate the desired operation to within the accuracy tolerance (if this is possible), with the number of gates in the circuit being close to minimal. We are working towards this goal by writing code that takes as an input an arbitrary quantum channel and provides a circuit corresponding to this channel as an output. Then, optimization algorithms can be applied to the found circuit, similar to the one introduced in [13]. Furthermore, we would expect that it is possible to speed up the optimization algorithm by using techniques from classical machine learning.

---

Prospects in Quantum Machine Learning
(Invited Talk)

Seth Lloyd
Massachusetts Institute of Technology

Quantum systems can generate statistics that cannot be generated by a classical system. This talk investigates the question of whether quantum systems can also recognize and classify statistics that cannot be recognized classically. Algorithms and technologies for quantum machine learning are presented and prospects for the field are discussed.
Knots, Computation and Quantum Physics
(Invited Talk)

Jiannis Pachos
University of Leeds, UK

Combining physics, mathematics and computer science, topological quantum computation is a rapidly expanding field of research focused on the exploration of quantum evolutions that are resilient to errors. In this talk I will present a variety of different topics starting from characterising knot invariants, proposing their quantum simulation with exotic quasi-particles called anyons and finally identifying the properties of strongly correlated systems that give rise to anyons.
Towards Quantum Machine Learning with Tensor Networks

William Huggins\textsuperscript{1}, Miles Stoudenmire\textsuperscript{2}

\textsuperscript{1}University of California, Berkeley, Berkeley, CA 94720, USA
\textsuperscript{2}Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

September 23, 2017

Abstract

We present the results of some experiments using tree tensor networks for supervised learning and share some of the lessons learned along the way. We highlight how our approach can inform future efforts to build hybrid quantum-classical algorithms for classification. Finally, we point out that the expressive power of the models we have considered would benefit from implementation on a quantum computer and draw connections to recent results regarding the expressivity of classical deep networks.

1 Introduction

In the near future small quantum computers with a limited size and accessible gate depth will be a reality. Even with their imperfections these devices are expected to be capable of performing calculations inaccessible to the largest classical computers \cite{1}. Meanwhile, there are a suite of proposals for quantum algorithms to accelerate various machine learning tasks \cite{2}–\cite{5}, but few of these approaches seem suitable for use on the relatively primitive machines that will soon be available. Hybrid quantum-classical algorithms provide a natural use case for these early generation devices, and the exploration of such algorithms dovetails nicely with the empirically driven experimentation that has been a hallmark of the recent explosion in deep learning. This work seeks to help lay the foundation for the coming experimental era of quantum machine learning by further exploring the use of tensor networks, a class of tensor decompositions that includes all quantum circuits, as tools for supervised classification.

Tensor network methods were originally developed in the context of condensed matter physics as tools for approximating the ground states of many-body Hamiltonians \cite{6}, \cite{7} using collections of low rank tensors along with a prescription for contracting them together. A combination of rigorous proofs \cite{8}, and strong numerical evidence \cite{9}, has shown that it is possible to tune a tensor network to represent the natural correlations of such states by tailoring the structure of the underlying graph of contractions to suit the problem. More recently, these tools have been applied to several problems in machine learning: Decompositions inspired by Matrix Product State \cite{7} wavefunctions (known as Tensor Trains in the mathematics community \cite{10}) have been used for image classification \cite{11}, recommendation systems \cite{12}, feature extraction \cite{13}, and other tasks. Furthermore, a connection has been shown between a restricted class of convolutional neural networks, convolutional arithmetic circuits with non-overlapping windows, and tree-like tensor networks \cite{14}.

Several of these works \cite{11}, \cite{12}, \cite{14} employ the same basic approach of lifting data into a high dimensional space by mapping it to a tensor product of simple feature vectors. After this initial non-linear transformation they use a tensor network to efficiently apply a linear operator in the expanded space. This is an approach that could be directly ported
over for use on a near term quantum device and is reminiscent of, but distinct from, the approach described in Ref. [5], where classical inputs correspond exclusively to computational basis states. Two questions naturally arise from these observations: Firstly, can this class of methods achieve results competitive with the current state-of-the-art in machine learning, and secondly, is there a benefit to using a quantum computer for these calculations? We now turn to some preliminary results that aim to address the first question.

2 Classification with Tree Tensor Networks

Building on the work in Ref. [11] we implemented a tree tensor network classifier and trained it on the MNIST handwritten digit dataset [15] using a generalization of their physics inspired sweeping algorithm [6] (a procedure which can be seen as a type of block coordinate descent with some additional desirable features). We prepare the inputs to our model by transforming an initial vector of normalized greyscale pixel values \( \vec{x} \) to the exponentially larger tensor product:

\[
g(\vec{x}) := \frac{1}{\sqrt{3}} \left[ \begin{array}{c} \cos \left( \frac{\pi}{2} x_0 \right) \\ \sin \left( \frac{\pi}{2} x_0 \right) \end{array} \right] \otimes \frac{1}{\sqrt{3}} \left[ \begin{array}{c} \cos \left( \frac{\pi}{2} x_1 \right) \\ \sin \left( \frac{\pi}{2} x_1 \right) \end{array} \right] \otimes \ldots \otimes \frac{1}{\sqrt{3}} \left[ \begin{array}{c} \cos \left( \frac{\pi}{2} x_{n-1} \right) \\ \sin \left( \frac{\pi}{2} x_{n-1} \right) \end{array} \right]
\]

Subsequently, we apply a series of linear maps in this expanded space, each acting on a pair of vectors. By iteratively coarse graining the input as depicted in Fig. 2 we generate a vector which we can transform into a probability distribution over the labels.

We have successfully combined the techniques of Ref. [11] with a mini-batch gradient descent training procedure, the dropout strategy of Ref. [12], and a regularization penalty inspired by the Von-Neumann entanglement entropy, allowing us to achieve a 97.9% accuracy on the MNIST test set with a relatively modest bond dimension (dimension of the output of our coarse-graining maps) of 10. Benchmarks of the model with more parameters and on more challenging datasets are currently in progress, although we expect that we’ll need more sophisticated feature maps and initialization strategies to achieve the best possible performance.

3 Going Forward

Our work thus far has been concerned with advancing the current state of the art for algorithms which act linearly on a tensor product of easy to construct feature vectors. The operations implemented by our model are not restricted to be unitary but it would be straightforward to approximate them by a combination of rotations and projections [16], or just constrain them from the outset, to obtain a classically simulable quantum circuit. Of course, we are also interested in these models when the application of the classifying map is not classically tractable. It was recently shown that convolutional arithmetic circuits (ConvACs) without overlapping windows, which are closely related to the tree tensor networks we have considered, are exponentially less expressive than similar models with overlapping windows [17], which have more in common with traditional convolutional neural networks but act non-linearly in the high dimensional feature space and so can not be interpreted as tensor networks. Fortunately, it is clear that even relatively shallow quantum circuits offer the same exponential advantage in expressivity that Ref. [17] found for overlapping ConvACs.

Whether or not the extra power that quantum classifying functions offer over classically tractable linear functions will ultimately useful is an open question. However, it seems clear that a better understanding of these approaches, and their practical limitations, will provide a strong foundation upon which to build a program of experimentation on tomorrow’s quantum devices.
References


Quantum Walk Neural Networks

Stefan Dernbach, Arman Mohseni-Kabir, Don Towsley
University of Massachusetts
Amherst, MA, USA
{Dernbach, Arman, Towsley}@cs.umass.edu

Siddharth Pal
Raytheon BBN Technologies
Cambridge, MA, USA
Siddharth.Pal@raytheon.com

INTRODUCTION

In recent years, neural network approaches have performed extremely well in diverse fields ranging from image recognition and classification [1] to natural language processing [2]. A key reason for the recent successes of neural network methods in the aforementioned domains is due to the statistical properties of the data in these domains, namely stationarity and constitutionality [3]. However, when the data in question is graph-structured, neural network techniques need to be adapted because the aforementioned properties might not hold. Despite the technical difficulties, new neural networks that operate on graph structured data have sprung into prominence [3], [4], [5]. These approaches use the graph Laplacian, random walk matrix, or their corresponding spectra as an operator to diffuse information across the graph and then learn a task-specific weighting on the diffused information.

In this work, we propose quantum walk neural networks (QWNN), a new graph neural network architecture based on quantum random walks. A quantum random walk differs from a classical random walk in that the walker’s state is expressed as a superposition rather than a probability distribution. Additionally, a coin operator acts on the walker at each step in the walk. Unlike previous graph neural networks, our approach uses this coin to directly learn a diffusion operator. We show that our quantum walk based neural network approach obtains competitive results when compared to other graph neural network approaches, suggesting that quantum techniques should be investigated further in the domain of graph-structured data.

QUANTUM WALKS ON ARBITRARY GRAPHS

Quantum random walks are the quantum parallel to classical random walks on a graph. While a classical walker is modeled by a probability distribution over positions in a graph, a quantum walker is described by a superposition over position states. We utilize the form of a discrete time quantum walk on a general graph as outlined in [6]. Given a graph $G = (V, E)$, we define a Hilbert space spanned by state $|v\rangle$ where $v \in V$. Also, we define $H_a$, the coin space, as an auxiliary Hilbert space of dimension $d_{\text{max}}$ spanned by the basis states $\{|i\rangle | i = 1, \ldots, d_{\text{max}}\}$, where $d_{\text{max}}$ is the maximum degree of the graph. We make the graph d-regular by adding self-loops to the vertices that have degree less than $d_{\text{max}}$. These states form the spin directions of a walker at vertex $v$. A single step in the quantum random walk consists of first applying a coin operator that transforms the coin state of a vertex, $C|v, e\rangle$. This coin operator is unitary and must treat all edges adjacent to a vertex equally. The Grover diffusion operator is the only nontrivial, real valued operator fitting these conditions.

After applying the coin operator, a shift operator swaps the states of two vertices connected by an edge using the following shift operation:

$$S|u, v, A_{uv}\rangle = \begin{cases} |u, v, A_{uv}\rangle & A_{uv} = 0 \\ |v, u, A_{uv}\rangle & A_{uv} = 1 \end{cases}$$

Where $A$ is the adjacency matrix of the graph. We define one step of the discrete quantum walk on graph $G$ as: $U = S(C \otimes I)$. If $|\Psi_0\rangle$ is the initial state of the quantum walker on $G$ then after $t$ time steps the state of the walker is described by: $|\Psi_t\rangle = U^t|\Psi_0\rangle$.

QUANTUM WALK NEURAL NETWORKS

Our new neural network architecture learns a quantum random walk on a graph by means of learning the coin operator. Our network then uses this learned quantum random walk to form a diffusion operator to act on the input data. Given a tensor $\Phi$ representing each walkers superposition, a coin matrix $C$ and a shift operator $S$, a quantum walk neural network takes in features $X$ and outputs diffused features $Y$. For the first QWNN layer in a network, we initialize $\Phi$ with a unique walker at each node in the graph and equal spin along each edge. For subsequent layers, the tensor $\Phi$ can additionally be passed forward with the feature matrix $Y$ to continue the walk. The method is given in Algorithm 1. The notation $a \cdot b$ denotes the inner product between tensors $a$ and $b$, the operator $a : b$ is the inner product over two dimensions, and $a \ast b$ is an elementwise product.

Algorithm 1 QWNN Forward Pass

given Superpositions $\Phi$, Coin $C$, Shift $S$
input Features $X$
for $t = 1$ to $T$
do
for All nodes $j$
do
$\Phi^{(t)}_j \leftarrow \Phi^{(t-1)}_j : C_j$,
end for
$\Phi^{(t)} \leftarrow \Phi^{(t)} : S$
end for
$P \leftarrow \sum_k \Phi^{(T)}_k \ast \Phi^{(T)}_k$ $Y \leftarrow P \cdot X$
return $Y, \Phi$
The work in [6] uses Grover’s diffusional operator for the coin operator because it is the only nontrivial, real-valued transform that is unitary and treats all edges connected to a vertex identically. The first requirement guarantees that location probabilities of the walker always sum to 1. The second restriction is added in order to avoid edge ordering affecting the walk. In the QWNN we relax these conditions to allow for learning biased coin operators. We initialize each coin to Grover’s diffusion operator. Then, during training the coin operators are adjusted by backpropagating the error signal through each layer without the above restrictions.

EXPERIMENTS

We demonstrate the use of our network in learning to predict daily temperatures. Our data consists of a years worth of daily high temperatures from 409 locations across the United Sates in 2009 [7].

We form a nearest neighbors graph from the stations’ longitudes and latitudes using 8 neighbors as this was empirically found to produce a fully connected graph (Fig.1). The temperature from each station on a single day is used to predict the following day’s temperatures. We form our neural network from a single quantum walk layer and vary the walk length. We also compare against a diffusion convolution neural network (DCNN) [5] while varying the number of hops in the network. The data is divided into thirds for training, validation, and testing. The mean absolute error (MAE) of the validation set over time is shown in Fig. 2. Fig. 3 gives the test results for the trained networks as well as a baseline predictor that predicts no change in the temperature. A QWNN with a walk length of 3 shows the best error out of all the methods.

CONCLUSIONS

The quantum random walk network demonstrates the power of quantum techniques for deep learning. With very few parameters to train, they demonstrate a large amount of predictive power as shown in the temperature forecasting experiment.

ACKNOWLEDGMENTS

This work was supported in part by Army Research Office Contract W911NF-12-1-0385 and ARL Cooperative Agreement W911NF-12-1-0385 and ARL Cooperative Agreement W911NF-12-1-0385 (the ARL Network Science CTA). This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

REFERENCES

Quantum machine learning for quantum anomaly detection

Nana Liu
Singapore University of Technology and Design, Singapore
Centre for Quantum Technologies, National University of Singapore
nana.liu@quantumlah.org

Patrick Rebentrost
Research Laboratory of Electronics,
Massachusetts Institute of Technology, USA
rebentr@mit.edu

Quantum computing has achieved success in finding algorithms that offer speed-ups to classical machine learning algorithms. This is especially important in the realm of big data. An important problem appearing in the presence of a large amount of input data is identifying unusual data patterns, or outliers, some coming from unreliable or unfamiliar sources. This is the subject of anomaly detection.

This problem is especially relevant when few examples of outliers may be available to develop a prior expectation. These outliers can be indicative of some unexpected phenomena emerging in a system that has never been before identified, like a faulty system or a malicious intruder. Anomaly detection is used for identifying data that deviate from ‘normal’ data patterns. Machine learning algorithms for anomaly detection in classical data finds diverse applications in many important areas like fraud detection, medical diagnoses, data cleaning and surveillance.

With the advent of quantum technologies, quantum states, in the form of quantum data, are prevalent in all forms of quantum computation, quantum simulation and quantum communication. Anomaly detection of quantum states is thus expected to be an important direction in the future of quantum information processing and communication, in particular, over the cloud or quantum internet. Since machine learning algorithms have proved successful for anomaly detection in classical data, a natural question arises if there exist quantum machine algorithms used for detecting anomalies in quantum systems. While classical anomaly detection techniques for quantum states can be used, they are only possible by first probing the classical descriptions of these states which require state tomography, requiring a large number of measurements. Thus, it would be advantageous to reduce these resource overheads by using quantum
In this work we discuss quantum machine learning methods applied to the detection of anomalies in quantum states themselves, which we call quantum anomaly detection. Given a training set of $M$ unknown quantum states, each of dimension $d$, the task is to use a quantum computer to detect outliers, occurring for example due to a faulty quantum device. Our schemes also do not require quantum random access memory, since the necessary data input is fully given by quantum states generated from quantum devices. In that sense, we present an instance of an algorithm for quantum learning.

We present two quantum algorithms for quantum anomaly detection: kernel principal component analysis (PCA) and one-class support vector machine (SVM). We show that pure state anomaly detection algorithms can be performed using resources logarithmic in $M$ and $d$. For mixed quantum states, we show how this is possible using resources logarithmic in $d$ only. We note this is also an exponential resource reduction compared to state tomography.
A quantum causal discovery algorithm

Christina Giarmatzi\textsuperscript{1} and Fabio Costa\textsuperscript{1}

\textsuperscript{1}Centre for Engineered Quantum Systems and Centre for Quantum Computation and Communication Technology, School of Mathematics and Physics, The University of Queensland, Brisbane QLD 4072, Australia

Causality plays a crucial role in science for the explanation and prediction of physical processes. The task of causal discovery, i.e. to extract causal relations between variables given some data, has been a common problem since the increasing amount of data from which we wish to extract some meaningful information. The field of machine learning has assisted on this problem by providing efficient methods for causal discovery. At the same time, other tasks of machine learning include causal discovery as an intermediate step.

In the recent years, the advancement of quantum techniques for computational tasks has given rise to quantum machine learning, which will inevitably face the problem of quantum causal discovery. This problem has attracted a lot of interest lately, from a foundational perspective. This is because, finding a causal model for a set of classical variables is a well-established task but what about quantum variables? Even the notion of a quantum causal model is ambiguous.

Here, we present an algorithm for quantum causal discovery [1]. We make use of a framework [2] that allows us to define causal models within quantum mechanics. In this framework, a causal model is represented by a Directed Acyclic Graph (DAG) in which the nodes and arrows represent quantum operations and quantum channels respectively. Making use of the process matrix formalism [3] it was found that a single DAG imposes linear constraints on the process matrix, given that the process is Markovian, namely that all relevant common causes are included in the process or that there are no latent variables not included in the process.

A process matrix describes an experimental scenario involving a set of parties, A, B, ... Each party is mathematically described by an input and an output Hilbert space and can perform an arbitrary operation (measurement, transformation, etc.). The probability that a set of outcomes is observed is given by the following formula

\[
P(s^A_1, s^B_1, \ldots | s^A_o, s^B_o, \ldots) = \frac{\text{Tr}(W^{A_1A_0B_1B_0\ldots}M^{A_1A_0}_{o|A} \otimes M^{B_1B_0}_{o|B} \otimes \ldots)}{\text{Tr}(W^{A_1A_0B_1B_0\ldots})}.\]

Here, \(s^X\) is the outcome of the party X given its setting \(s^X\), and \(M^{X_1X_0}\) is an operator that represents the operation of the party. (In fact it is a matrix isomorphic to a completely positive map from the input to the output system of each party, using a version of the Choi-Jamiolkowski (CJ) isomorphism.) \(W\) is the process matrix, that allows the parties to calculate their joint probabilities of their outcomes, given their operations. The superscripts on \(W\) and \(M\) denotes that each matrix lives on the tensor product of the Hilbert spaces of the input and output systems, i.e. \(X_I\) and \(X_O\) respectively, for the party \(X\).

With the above formula, the process matrix can be constructed from measured probabilities when the parties perform informationally complete operations. Our problem of causal discovery is then solved by extracting information encoded in the process matrix about the DAG. This is done by the observation that each element of the DAG (an edge between two parties, which party is first or last etc) corresponds to a particular linear constraint on the matrix.

For example, if the process involves three parties, labelled \(A, B, C\), with a causal order \(A \prec B \prec C\), then it will have the following (although unknown at first) form

\[
W^{A_1A_0B_1B_0C_1C_0} = \rho^{A_1} \otimes T^{A_0B_1} \otimes T^{B_0C_1} \otimes I^{C_0}\]

where \(\rho\) is some input state for party \(A\), \(T^{A_0B_1}\) is the Choi representation of a channel from (the output of) \(A\) to (the input of) \(B\), and similarly \(T^{B_0C_1}\) represents a channel from \(B\) to \(C\). Finally \(I^{C_0}\) is the identity matrix and represents the fact that \(C\) is last. We note that the process matrix will have this form only if the process is Markovian.

Our causal discovery algorithm takes advantage of the fact that the existence of each of these terms in \(W\), corresponds to a linear constraint on the process matrix. For example a channel from \(A\) to \(B\) exists if and only if the process matrix satisfies the following constraint

\[
\hat{I}^{A_0} \otimes \text{Tr}_{A_0}(T_{B_1}W) = \text{Tr}_{B_1}W, \]

where \(\hat{I}^{A_0} = \hat{I}^{A_0}/d_{A_0}\), and \(d_{A_0}\) is the dimension of the system \(A_0\). The constraint originates from the fact that the Choi representation of a CPTP map (a quantum channel) must satisfy the fact that upon tracing out the output of the channel, we obtain identity on the input of the channel. In a similar fashion we have a linear constraint for every term of the process matrix that is associated with an element in the DAG.

Therefore, we have a way that connects observational data (through the process matrix) to a DAG in the quantum case. However, reconstructing the DAG from a...
process matrix is an intricate and lengthy procedure, requiring a number of tests on the process matrix. One of these tests is to check if the process matrix satisfies a set of all possible linear constraints, that is all possible elements that the DAG might have. The set of the linear constraints that the process matrix satisfies, uniquely identifies a DAG, for a Markovian process. The algorithm we present automates this set of actions.

The algorithm takes as inputs: the number of parties, the dimensions of all input and output systems, if and how the output systems factorise into subsystems that are sent to different parties, and finally the process matrix. The code first establishes the consecutive sets, which provide information about the sets of parties that are causally independent and their causal order. Then it checks whether parties that belong to different sets are causally connected. For each such causal connection, the algorithm creates the channel, extracted from the process matrix, that represents this causal link. Finally, the code composes a matrix of the form \( (2) \) compatible with the found DAG. If this composed matrix equals to the process matrix, then the process is Markovian.

Hence, the algorithm outputs: 1) whether the process is Markovian or not, namely, whether all relevant common causes are included in the process, or else some causal relations are mediated through some external ‘memory’, and 2) (if the process is Markovian) a DAG describing all the causal relations between the parties, and which output subsystem (if any) of a party belongs to each causal arrow. Our algorithm provides a first step towards more general methods for causal discovery of quantum variables. This is a timely problem to be addressed as the increasing development of quantum technologies will require large, distributed quantum networks in which causal discovery will be a non-trivial and useful task.

Below is the output of the code, for an example of a process matrix regarding 4 parties, labelled 1, 2, 3, 4. Figure 1 shows the output of the code on the command window, which is information about the causal links between different systems of the parties. Note that the output system of each party could be decomposed into different output subsystems. An output subsystem to be ‘open’, means that there is no outgoing causal arrow from that system. Figure 2 shows the output DAG.

For more information, see Ref. [1] and for the implementation of the code, a manual and some examples of inputs to the code, see [4].

![Figure 1. Output of command window: Causal information regarding the consecutive sets, the open subsystems and the causal arrows. Primal_arrows refers the the causal arrows from successive consecutive sets and secondary_arrows refers to all the rest causal arrows. Time refers to the time that lapsed to evaluate the step just above.](image1)

![Figure 2. Output DAG: The DAG that the code outputs for the given example.](image2)

Progress in Quantum Reinforcement Learning
(Invited Talk)

Vedran Dunjko
Max Planck Institute of Quantum Optics, Germany

Recent work in quantum machine learning has demonstrated that quantum computers can offer dramatic improvements in various data analysis aspects of machine learning. However, machine learning, or broader, artificial intelligence machinery can do more than data analysis. Interactive learning methods, i.e. reinforcement learning, have driven some of the most thrilling recent technological and scientific trends: from the Go-playing AlphaGo, intelligent personal assistants (e.g., Siri, Amazon Alexa), to self-driving cars, and smart factories. Investigations into the advantages that quantum computers may bring about in such interactive learning scenarios have only recently produced first encouraging results. In this talk we will present the overview of this nascent topic of quantum reinforcement learning, including novel results suggesting that also in such more general learning scenarios, quantum computers may yield dramatic advantages.
Quantum Neural Networks: A Hamiltonian Complexity Approach

The goal of this work is to build a quantum neural network, train it on some actual data and then test its performance on a classification task. More precisely, the idea of this submission is to use Microsoft’s quantum simulator, LIQ\(U_i\), to construct local Hamiltonians that can encode trained classifier functions in their ground states, and which can be probed using simulated annealing. We propose a training scheme which is completely closed-off to the environment and which does not depend on external measurements, avoiding unnecessary decoherence during the annealing procedure. For a network of size \(n\), the trained network can be stored as a list of \(O(n)\) coupling strengths. We address the question of which interactions are most suitable for the classification task, developing a number of qubit-saving optimizations for the simulation. Furthermore, a small neural network to classify colors into red vs. blue is trained and tested.

Johannes Bausch — jkrb2@cam.ac.uk, DAMTP, University of Cambridge, UK — September 23, 2017.

Git repository at https://bitbucket.org/rumschuettel/liquidlearn.git

1 Introduction

The authors of [1] summarize past efforts to extend the notion of classical neural networks to the quantum world. This task is not straightforward: measurements have to be included to introduce nonlinearity, which interferes with the coherence of the model, or postselection is used to collapse a superposition of networks. A system of interacting quantum dots is suggested, capable of simulating network dynamics. However, experimentally, it is unclear how this is easier than building an analogue quantum computer.

Adiabatic quantum annealing [2] is the only area of experimental quantum computing claiming significant qubit counts to date. By tuning the couplings between qubits, one hopes to encode a computation into the ground state of the Hamiltonian. Cooling and measuring the system then reveals the computation output. The interactions, however, are typically limited to Ising-type couplings, and the physical interaction graph is fixed: logical interactions have to be crafted onto this underlying structure. Furthermore, the coupling strengths have to be hand-tuned to represent a problem of interest, or calculated on classical hardware.

Despite these limitations, specialized hardware such as the D-Wave—with their underlying Chimera interaction graph—has one significant advantage over a full-fledged analogue programmable quantum computer: we can build them today.

We introduce a new notion of quantum neural networks, based on this idea of quantum annealing on a fixed underlying graph, but addressed from a Hamiltonian complexity perspective. Furthermore, we propose a training algorithm which could be completely incorporated into such an annealing device, avoiding the need for any classical optimization or repeated measurement to introduce non-linearity, and reducing unnecessary decoherence. We construct a few such neural networks with LIQ\(U_i\)-Learn—an F# library based on LIQ\(U_i\) which can be used to build, train and test quantum classifiers—demonstrating that the quantum training algorithm works as intended. Finally, we address a few related questions, such as which interactions are most suitable for the network, and put it to the “real-world” classification task of distinguishing red and blue colors.

2 Techniques

We want to model a quantum neural network as a geometrically local hypergraph, where the edges represent interactions from a fixed set and hope that quantum effects—entanglement and long-range correlations—will step in to replace the lack of non-local connections in the interaction graph. Training will be done by varying the coupling strengths between adjacent qubits by replacing each interaction with a controlled one, and projecting onto the complement of a superposition over the training data on the external vertices: by measuring the control qubits in the ground state, one can deduce the optimal interaction strength. A schematic can be found in fig. 1.

More formally, we take a hypergraph \(G = (V, E)\) with a set of vertices \(V\) labelling qubits, and hyperedges \(E\) encoding interactions between them. Assign the qubit Hilbert space \(C^2\) to each vertex, such that the overall Hilbert space becomes \(\mathcal{H} = (C^2)^\otimes|V|\). For a local operator \(\mathbf{h}\) acting only on a subset of vertices \(U \subseteq V\) and as zero everywhere else, we write \(\mathbf{h}^{(U)}\). For every hyperedge, we are given a list of local interactions \((S_e)_{e \in E}\). The overall Hamiltonian on \(\mathcal{H}\) can then be written as

\[
H = \sum_{e \in E} \sum_{h \in S_e} a_{h,e} \mathbf{h}^{(e)},
\]

where the \(a_{h,e}\) are the coupling constants to train. We label some of the vertices of the hypergraph as output vertices \(V_o \subseteq V\), and call the complement \(V_o^c\) hidden.
red training colors
blue training colors

Figure 2: Training graph for 9 bit color classification problem. Each component encodes its color in 3 bits. We need 22 qubits to train the interactions. The right hand side shows ten randomly chosen red and blue shades used for training the classifier.

\[ \text{tr}_A(H) \text{ then denotes the Hamiltonian } H, \text{ where we traced out this hidden part of } H. \]

This setup allows us to define the following optimization problem.

**Definition 1 (Classification Problem)** Let \( L \) be a decision problem with instances of binary size \( n \). Given a hypergraph \( G \) with \( n \) output vertices, and a list of local interactions \( \{S_1\}_{E} \), find the optimal coupling constants such that the two discrete distributions \( D_{\text{YES}} := \{ |\text{tr}_A(H)| : l \in L_{\text{YES}} \} \) and analogously \( D_{\text{NO}} \) are maximally distinguishable.

Given \( G_{\text{YES/NO}} \), we can calculate soundness, defined as the fraction of \( l \in D_{\text{YES}} \) for which \( P(l \in G_{\text{YES}}) > P(l \in G_{\text{NO}}) \), as well as completeness, defined by flipping the roles of YES and NO. We further require the mean of \( G_{\text{YES}} \) to lie below the mean of \( G_{\text{NO}} \), in analogy to the hardness results summarized in the introduction. The definition deliberately leaves room for interpretation, i.e. what precisely maximally distinguishable means. For our purposes, we distinguish between the two distributions using a linear logistic regression.

### 3 Results

**A Real World Classification.** We would like to assess our setup with a “real-world” problem. Unfortunately, even the smallest examples from well-known machine learning datasets require way too many qubits as input to run on quantum simulators available today, so we have to construct a somewhat artificial dataset which works within the hardware and software limitations at hand.

We consider the task of classifying colors into the classes red and blue. Our training dataset will consist of a list of 9 bit, little-endian color codes in RGB order, e.g. 001 100 111, which would represent a light blue tone (little red, half green, full blue). This is an interesting, non-trivial task, as there are a lot of edge cases, as well as some colors which are neither considered red or blue. So let us see what our quantum classifier thinks.

The graph we will train this example on can be seen in fig. 2. As interactions, we use 2-local projectors on computational basis states. Ten red and blue shades are chosen randomly to train the network, and we test it on the full possible 512 color spectrum available with 9 bits. The output is plotted in fig. 3, sorted from most red to most blue. The error bar shows the standard deviation as returned by the test measurement utilizing LIQ/Ui’s Spin.Energy.Expectation.

**Benchmarking Interactions.** Which type of interaction is most suited for a quantum classifying Hamiltonian, in the sense that it requires few qubits to train while giving a good separation between the YES and NO instances?

Pauli matrices give a full basis for Hermitian matrices, as we can expand any Hermitian as a linear combination of Pauli product matrices. Ising-type interactions are relevant, as they play a fundamental role in many real-world physical systems and are the interactions used in a D-Wave, so it is interesting to see how they compare to others. Heisenberg interactions have been proven to be QMA-complete in [10], and we further analyse random matrices as a benchmark for the others.

In order to quantify how these interaction sets fare, we try to train all possible datasets on small quantum networks. The intuition is that if an interaction type is able to correctly train a lot of combinations on a small scale, assembling a bigger graph from these interaction types results in a more space-efficient network.

A priori, it is not clear which interactions will perform best in a specific scenario. The underlying graph plays a big role, as well as overfitting, which the Pauli matrices seem to suffer from: their marginally better energy separation as compared to other interaction types might be a bad tradeoff in terms of runtime. Surprisingly, the randomly chosen sparse matrices fare better than expected, and Heisenberg and Ising interactions are usually outperformed.
References


Quantum Entanglement Simulators Inspired by Tensor Network

Shi-Ju Ran

ICFO  The Institute of Photonic Sciences, Spain

Abstract

Tensor network (TN), a young mathematical tool of high vitality and great potential, has been undergoing extremely rapid developments in the last two decades. We report a scheme of constructing the few-body models dubbed quantum entanglement simulators (QES’s) that can be easily accessed by theoretical or experimental means. We show that a QES accurately captures the ground-state properties of infinite many-body systems in any dimensions. The general idea is to embed a small bulk of the infinite model in an “entanglement bath” so that the many-body effects can be faithfully mimicked. The approach we propose is efficient, simple, flexible, sign-problem-free, and it directly accesses the thermodynamic limit. The numerical results of the spin models on honeycomb and simple cubic lattices show that the ground-state properties including quantum phase transitions and the critical behaviors are accurately captured by only $O(10)$ physical and bath sites. Moreover, since the few-body Hamiltonian only contains local interactions among a handful of sites, our work provides new ways of studying the many-body phenomena in the infinite strongly-correlated systems by mimicking them in the few-body experiments using cold atoms/ions, or developing novel quantum devices by utilizing the many-body features.
Measuring Entanglement Negativity with Neural Network Estimators

Johnnie Gray,1 Leonardo Banchi,1 Abolfazl Bayat,1 and Sougato Bose1

1Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom

(Dated: September 16, 2017)

Entanglement not only plays a crucial role in quantum technologies, but is key to our understanding of quantum correlations in many-body systems. However, in an experiment, the only way of measuring entanglement in a generic mixed state is through reconstructive quantum tomography, requiring an exponential number of measurements in the system size. Here, we propose an operational scheme to measure the entanglement – as given by the negativity – between arbitrary subsystems of size \(N_A\) and \(N_B\), with \(O(N_A + N_B)\) measurements, and without any prior knowledge of the state. We show that neural networks can be trained to accurately extract the negativity from a few moments of the partially transposed density matrix, which can be experimentally measured. Our procedure will allow entanglement measurements in a wide variety of systems, including strongly interacting many body systems in both equilibrium and non-equilibrium regimes.

Entanglement is a key quantum property in multiple fields. For quantum technologies, it is a resource that allows enhanced sensing, faster computation and new paradigms of communication. In condensed matter, entanglement is essential for understanding the structure of strongly correlated many-body systems and has allowed the development of efficient classical algorithms for simulating many-body systems. It is thus crucial to be able to measure entanglement for mixed states in an experimental setting. While for pure states bipartite entanglement is uniquely defined by the entropy of the subsystems, this ceases to quantify entanglement for the more generic case of mixed states. Here, the landscape of bipartite entanglement measures is far more complex, and aside from isolated special cases such as two qubit states and bosonic Gaussian states, only the (logarithmic) negativity is a computationally tractable quantity. Nonetheless, there is no state-independent observable that can measure the logarithmic negativity, and thus it is impossible to experimentally measure it without reconstructing the state using quantum tomography – a generically demanding process with an exponential number of measurements in the system size. Here, we put forward a practical proposal for accurately estimating the logarithmic negativity in a completely general and realistic setting, using an efficient number of measurements – scaling polynomially with system size. Our method is based on measuring a finite number of moments of a partially transposed density matrix (see [1]), from which we extract the entanglement negativity using neural networks. This represents a new front in the emerging arena of classical machine learning methods applied to quantum estimation and measurement problems. Moreover, our method is experimentally feasible, since the individual building blocks have already been demonstrated in solid state and cold atoms.

**Logarithmic Negativity.** Logarithmic negativity for a generic mixed state \(\rho_{AB}\) quantifies the quantum entanglement between subsystems \(A\) and \(B\) – the mixedness of \(\rho_{AB}\) arises from possible entanglement with a third system \(C\). It is defined as: \(\mathcal{E} = \log_2 |\rho_{AB}^{T_A}| - \log_2 |\rho_{AB}| = \log_2 \sum_k |\lambda_k|\), where \(|\cdot|\) denotes the trace norm, \(\rho_{AB}^{T_A}\) is the partial transpose with respect to subsystem \(X\), and \(\{\lambda_k\}\) are the eigenvalues of \(\rho_{AB}^{T_A}\). Because of the non-trivial dependence of \(\mathcal{E}\) on \(\rho_{AB}\), there is no state-independent observable that can measure it — one generally needs to reconstruct the full density matrix via state tomography. The \(\{\lambda_k\}\) are the roots of the characteristic polynomial, \(P(\lambda) = \det(\lambda I - \rho_{AB}) = \sum_n c_n \lambda^n\), where each coefficient \(c_n\) is a polynomial function of the partially transposed moments: \(\mu_m = \text{Tr}[(\rho_{AB}^{T_A})^m] = \sum_k \lambda_k^m\). In this way, full information about the spectrum \(\{\lambda_k\}\) is contained in \(\{\mu_m\}\), with the first few moments carrying the most significance. Indeed, since \(-\frac{1}{2} \leq \lambda_k \leq 1\) for all \(k\) and \(\sum_k \lambda_k = 1\), generically, the magnitude of the moments quickly decreases with \(m\). We will show that \(\mathcal{E}\) can be accurately estimated using only \(\{\mu_m : m \leq M\}\), with \(M\) as low as 3. Note that \(\mu_0\) is simply the dimension of the systems Hilbert space, while \(\mu_1 = 1\) in all cases. Additionally, it can be easily shown that \(\mu_2\) is equal to the purity of the state \(\text{Tr}[\rho_{AB}^2]\), and as such, \(M \geq 3\) is needed to extract any information about \(\mathcal{E}\). As discussed in [1], the moments \(\mu_m\) can be measured efficiently in different physical systems using \(m\) copies of the original systems, and then performing SWAP measurements on neighbouring copies. The total number of measurements required is \(O[\mu_m(N_A + N_B)]\).

**Estimating Entanglement.** The relationship between the moments \(\{\mu_m\}\) and the negativity \(\mathcal{E}\) is inherently non-linear. Machine learning has recently emerged as a key tool for modelling an unknown non-linear relationship between sets of data. In the supervised learning paradigm, one trains a model with a set of known inputs and their corresponding outputs. Once trained, the model can then be used to predict the unknown output of new input data. Here, we take the moments \(\mu_m\) as the input and the logarithmic negativity \(\mathcal{E}\) as the output. Training is performed by taking a large set of states for which \(\mu_m\) and \(\mathcal{E}\) can be computed on a classical computer. This model can then be used to predict \(\mathcal{E}\) from a set of experimentally measured moments. The experimental system under study motivates the choice of which training states to use, so that they share, for example, similar entanglement features. Among the
most successful machine learning algorithms for non-linear regression are supervised vector machines, random decision forests, and deep neural networks. However, we have found that using the same training set for each, neural networks are superior when it comes to predicting logarithmic negativity for a wide range of states beyond the training set.

From an entanglement perspective, relevant states in condensed matter physics can be classified as either area-law, with the rest is proportional to the number of qubits along their boundary. In the second, this entanglement is instead proportional to the number of qubits along their boundary. In the first case, the entanglement of a subsystem evolves in time as \( |\Psi(t)\rangle = e^{-\frac{i}{\hbar}Ht} |\Psi(0)\rangle \). As the chain unitarily evolves in time as \( |\Psi(t)\rangle = e^{-\frac{i}{\hbar}Ht} |\Psi(0)\rangle \), it becomes entangled. and logarithmic negativity displays oscillatory dynamics. We plot the logarithmic negativity, \( \mathcal{E} \), and its approximations, using machine learning (\( M = 3 \)) and a Chebyshev expansion (\( M = 10 \)), as a function of time \( Jt \). A wide range of system sizes with different partitions is shown here: (a) \( N = 5, N_A = N_B = 1 \); (b) \( N = 8, N_A = N_B = 2 \). (c) \( N = 11, N_A = 3, N_B = 5 \); (d) \( N = 20, N_A = N_B = 5 \).

**Numerical Results.** In Fig. 1 we check the performance of neural network estimators, obtained using Hyperopt and Keras packages to find the optimal network structure, including the number of hidden layers – for more details, see [1]. The model is trained using the different random states described above, while predictions are made for both random states and physical states (thus mimicking a physical experiment). Here, the machine learning approximation, using only \( M = 3 \) copies, significantly outperforms analytical methods based on the Chebyshev approximations (see [1]), even when the latter uses more copies. It is remarkable that despite being trained on a arbitrary set of random states with no knowledge of the underlying physical system, the evolution of \( \mathcal{E} \) is accurately captured by the neural network estimator for all partitions and times, with as few as \( M = 3 \) copies.

Quantum state engineering using one-dimensional discrete-time quantum walks

Luca Innocenti1, Helena Majury 1,2, Taira Giordani2, Nicolò Spagnolo3, Fabio Sciarrino3,4, Mauro Paternostro1,5, and Alessandro Ferraro1

1Centre for Theoretical Atomic, Molecular, and Optical, School of Mathematics and Physics, Queens University Belfast, BT7 1NN Belfast, United Kingdom
2Centre for Secure Information Technologies, Queens University, Belfast BT7 1NN Belfast, United Kingdom
3Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy
4Consiglio Nazionale delle Ricerche, Istituto dei sistemi Complessi (CNR-ISC), Via dei Taurini 19, 00185 Roma, Italy
5Laboratoire Kastler Brossel, ENS-PSL Research University, 24 rue Lhomond, F-75005 Paris, France

Quantum state preparation in large-size Hilbert space is a fundamental step in developing quantum technologies. The applications of high-dimensional quantum state range from quantum computation in general to quantum error correction. Consequently exploiting new quantum state engineering protocols is of crucial interest. Many strategies have been theoretically proposed and experimentally realized in quantum optical systems using a variety of degrees of freedom, from time-energy to polarization, path, orbital angular momentum and frequency. In this work we have developed a scheme based on Discrete-Time Quantum Walks on a line (DTQW) and a possible experimental implementation with linear optics and angular momenta of light.

A quantum walk, the quantum counterpart of classical random walks, consists of a d-dimensional system usually named as walker’s site, with an inner 2- dimensional degree of freedom, referred to as the coin state of the walker [1]. It has been demonstrated that quantum walks are universal models for quantum computation [2], and, in particular, allow for efficient implementations of quantum search algorithms [3]. A single step of DTQW evolution is made up of coin flipping in which an unitary coin operator C is applied to the coin degree of freedom of the walker, and a walking step, in which the walker’s site is changed conditionally to the state of the coin, with a shift operator S. Formally, given an initial state of the walker $|\Psi_{in}\rangle \equiv \sum_{k=1}^{d} \sum_{\alpha=\{\uparrow, \downarrow\}} u_{k,\alpha} |k\rangle \otimes |\alpha\rangle$, after a step the state becomes

$$W_{C} |\Psi_{in}\rangle \equiv S C |\Psi_{in}\rangle = \sum_{k=1}^{d} \sum_{\alpha=\{\uparrow, \downarrow\}} u_{k,\alpha} S \left( |k\rangle \otimes C |\alpha\rangle \right),$$

(1)

where we defined the step operator $W_{C}$ as the combined action of a coin flip and a controlled shift, C is the coin operator, and S is the controlled-shift operation defined as

$$S = \sum_{k} \left( |k\rangle\langle k| \otimes |\uparrow\rangle\langle \uparrow| + |k\rangle\langle k| \otimes |\downarrow\rangle\langle \downarrow| \right),$$

(2)

The amplitudes $u_{k,\alpha}^{(n)}$ of resulting state $|\Psi^{(n)}\rangle$ after n steps of a quantum walk must satisfy a set of constrains, in particular a series of $2n - 2$ real conditions. It follows that space of reachable states by a DTQW has dimension $2n$. In this way these constraints identify the set of the possible high-dimensional quantum state that are the output of a quantum walk’s routine. This $2n$-dimensional Hilbert space $H^{2n}$ can be covered allowing the coin operator to change at every step. Conversely, given the constrains founded in this work, not only every state that is the result of a quantum walk evolution satisfies them, but also for every state $|\Psi\rangle \in H^{2n}$ that satisfies the conditions there is a set of n coin operators $\{C_{i}\}_{i=1}^{n}$ and an initial state $|\Psi_{in}\rangle$ such that the state $|\Psi\rangle$ can be also obtained trough a DTQW with steps operators $\{W_{C_{i}}\}_{i=1}^{n}$. The collection of coin operators needed to reach a certain quantum state is efficiently computable. The target state of the protocol is then obtained after
Let us propose an experimental implementation in which it is possible to test the quantum state engineering protocol proposed above. A DTQW can be realized with linear optics, in particular, encoding the walker’s site and the coin, respectively, in Orbital Angular Momentum (OAM) and Spin Angular Momentum (SAM) of a photon [4]. The complete scheme is shown in Fig.1. The walker’s degree of freedom is described by the quantum number $m \in \mathbb{Z}$, eigenvalue of the OAM along the propagation axis, and the coin state is encoded in left $|L\rangle$ and right $|R\rangle$ polarization states of the photon. The encoding in polarization allows us to implement any arbitrary coin operations, propagating the photon through quarter (QWP) and half (HWP) wave plates. The shift operators are instead realized with $q$-plates (QP): a birefringent liquid-crystal medium that rises and lowers the value of $m$, according to their topological charge $q$, conditionally to the polarization state. A QP changes the wavefront transverse profile of the photon without deflections [5]. Therefore DTQW’s steps are made up of consecutive optical units composed of wave plates (coin operators) and a QP (shift operators). The number of optical elements scales polynomially with the number of steps, making this scheme scalable.

In conclusion, allowing the coin operator to vary at each step, a set of conditions to identify the reachable states by a coined quantum walk evolution has been provided. We proposed a possible experimental realization of the scheme with linear optics, using orbital and spin angular momentum of a photon to encode spatial and coin degrees of freedom of the walker. Thanks to the richness offered by quantum walks dynamics and the easiness to realize them on a linear optical platform, the approach proposed to quantum state preparation should facilitate the engineering of high-dimensional quantum states in a wide range of quantum information protocols.
References

Quantum neuron  
(extended abstract)

Yudong Cao1,∗ Gian Giacomo Guerreschi2 Alán Aspuru-Guzik1

1Harvard University, Cambridge, MA 02138, USA  
2Intel Corporation, Santa Clara, CA 95054, USA

Efforts for combining the unique features of quantum computing and neural computing have motivated the search for Quantum Neural Networks (QNN). The fundamental building block of any neural network is the artificial neuron, which combines weighted signals from its input, applies a non-linear filter (such as a threshold or sigmoid function) to the combined signal and outputs the filtered signal to other neurons. Here we present a construction of quantum circuits which captures the core features of an artificial neuron. Our circuit construction, which we call a quantum neuron, allows the input signal to be a quantum superposition, thus taking advantage of the unique feature of quantum mechanics.

An overview of the construction is shown in Figure 1. Here we derive an analogy from classical artificial neuron (Figure 1a) where the state of the neuron is a real number $x \in [-1, 1]$. In our quantum neuron, we use the state of a qubit to represent the state of the neuron. Extremal values $x = -1$ and $x = 1$ in the classical case translate to quantum states $|0\rangle$ and $|1\rangle$ respectively, and the intermediate values between $-1$ and $1$ in the classical case correspond to states of the form $R_y(\theta)|0\rangle$ where $R_y(\theta) = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} Y$ and the value $\theta$ lies between $0$ and $\pi$. Therefore one could consider preparing a qubit in the state $R_y(\theta)|0\rangle$ as an action of generating a signal of strength $\theta$. If the classical neuron is regarded as a device which transforms an input signal $x$ to some output signal $\sigma(x)$ with $\sigma$ being some non-linear function such as threshold or sigmoid function, quantum mechanically we accomplish such transformation (Figure 1b) using quantum circuits known as repeat-until-success circuits (Figure 1c and d). The circuit in Figure 1c accomplishes the non-linear transformation from input signal $\theta$ to an output $q(\theta)$. More iterations of the circuit construction could yield $q(q(\ldots q(\theta))))$, which has sharper non-linearity (Figure 1d and also see Wiebe and Kluchnikov1).

We use two examples to show how to use our quantum neuron to build quantum neural networks. The first example is the feedforward neural network. For a given training set consisting of input data $\{\vec{x}_i\}_{i=1}^n$ and output $\{\vec{y}_i\}_{i=1}^n$, we prepare the initial state which is a superposition of $|\vec{x}_i\rangle|\vec{y}_i\rangle$ states. Then we propagate the input state through a network of quantum neurons and use the final states of the output neurons as the “prediction” of the quantum neural network. Finally we evaluate the accuracy of the “prediction” by pairing up each output neuron with the corresponding qubit in the register holding $|\vec{y}_i\rangle$ states and measuring the $\langle ZZ\rangle$ expectation on all pairs. We numerically demonstrate training with superposition of training data and yield both training and testing accuracy above 99% for XOR and parity functions (Figure 2). The potential quantum advantage in our construction is that our construction allows training over all possible inputs, instead of a selected training set as is the case for classical neural networks.

For users of artificial intelligence who intend to use deep neural networks for extracting patterns from training data, our quantum neural network can enable a novel form of deep learning that deliver higher training throughput and efficiency. Unlike current deep neural networks, the quantum neural networks provide ability to train neural networks using a superposition of multiple data at once.

The second example is the Hopfield network, where we show attractor dynamics given a superposition of network initial states. For both types of neural networks we provide rigorous connections to their classical counterparts. To conclude, we argue that our construction is the first one which satisfies all of the major criteria proposed by Schuld et al2 for a reasonable construction of quantum neural networks.


∗yudongcao@fas.harvard.edu
Figure 1: Schematic design of the quantum neuron model. (a) The classical neuron where the inputs are \( \{ x_i \}_{i=1}^n \) and the output \( a = \sigma(w_1 x_1 + \cdots + w_n x_n + b) \), with the activation function \( \sigma \) being a sigmoid or step function. (b) The overall behaviour of a quantum neuron. The states of the input qubit and the output qubit are in their own Bloch sphere. States \(|0\rangle\) and \(|1\rangle\) are aligned along the vertical axis, respectively pointing down and up. The function \( q \) is shown in subplot (d). The notation \( \rightarrow \) means “A \( \rightarrow \) B iff given the ability to realize A one could realize B”. The input state is assumed to be prepared by some external method, possibly controlled by other quantum neurons. (c) Repeat-until-success circuit for realizing rotation with an angle \( q(\varphi) = \arctan^2 \varphi \). (d) Nonlinear function \( q(x) \) and \( q^{\circ k}(x) \). Observe as \( k \) increases the nonlinearity is sharper.

Figure 2: Numerical results for optimizing feedforward neural network. Here we consider (a) XOR network and (b) Parity network. Different colors represent runs with different initial guesses for weights and biases. Networks of blue circles are quantum neurons (qubits) in the network. The red circle is the qubit holding the training label with which we compute \( \langle ZZ \rangle \) together with the output neuron (qubit) for evaluating training accuracy.
Quantum Simulation of Complex Hamiltonians With Directionally Unbiased Linear-Optical Multiports

Alexander V. Sergienko, David S. Simon, Shuto Osawa, and Casey A. Fitzpatrick

Abstract: The directionally unbiased optical multiport is introduced, in which photons may reflect back out the input direction. Quantum walks on arrays of such multiports allow the simulation of a range of complex quantum Hamiltonians.

OCIS codes: (270.0270) Quantum Optics; (270.5585) Quantum Information and processing

Quantum computers have been shown to be capable of performing certain kinds of tasks exponentially faster than classical computers. As a result, an enormous amount of effort has gone into their development. Although advances have been made, the ultimate goal of a large-scale programmable, general-purpose quantum computer still seems to be a relatively long way off. Therefore it is useful to consider the more easily attainable possibility of special-purpose quantum computers designed to carry out specific tasks. In particular, one might consider returning to Feynman's original motivation for discussing quantum computers [1]: using simple quantum systems to simulate the behavior of other physical systems.

Recently, a generalization of the standard linear optical multiport was proposed [2] which allows photons to reverse direction, thus transcending feed-forward linear optics by providing a linear-optical scattering vertex for quantum walks on arbitrary graph structures. The devices impose group structures on two-photon entangled Bell states and act as universal Bell state processors to implement probabilistic quantum gates acting on state symmetries. These multiports allow optical scattering experiments to be carried out on arbitrary undirected graphs via linear optics, and raise the possibility of linear optical information processing using group structures formed by optical qudit states.

Symmetry has long been a guiding principle in physics. Central to linear optics is the beam splitter (BS), which has reflection symmetry as well as a time-reversal symmetry in which a photon may either enter or leave any of the four ports. However, the BS is asymmetric in another sense.
Once a photon enters a port, it may not leave from the same port: a photon in one port breaks the symmetry. We call this inability to reverse direction directional bias. Here we introduce a linear-optical arrangement that restores the symmetry: it is directionally unbiased (see Fig. 1). This arrangement, which can be thought of from multiple viewpoints (as a directionally unbiased multiport BS, as a resonant cavity with three or more exit directions, or as a scattering vertex for an undirected graph), is generalizable to any number of ports \( n = 3 \); we focus on the simplest case of \( n = 3 \). Despite its simplicity, the high degree of symmetry and the ability of photons to reverse direction lead to interesting properties that allow it to be used as the basis for a variety of applications related to quantum information processing [2].

Quantum walks on arrays of such multiports will allow the simulation of a range of discrete-time Hamiltonians representing complex physical systems [3].

Several practical approaches to quantum simulation of complex Hamiltonians are proposed [3], including cases where physical systems with both spatial and internal degrees of freedom as well as topological states are simulated [4]. The simulation is carried out in a scalable manner, using only linear optics, and can be generalized to higher dimensional systems in a straightforward manner, thus offering a concrete instantiation of graphical models of discrete time quantum systems. Because input ports also double as output ports, there is substantial savings of resources compared to traditional feed-forward quantum walk networks carrying out similar functions.

References

Quantum Error Correction with Recurrent Neural Networks

P. Baireuther, T. E. O’Brien, B. Tarasinski and C. W. J. Beenakker

Instituut-Lorentz, Universiteit Leiden, Netherlands
Delft University of Technology, Netherlands

1 Extended Abstract

In quantum computation one of the key challenges is to build fault-tolerant logical qubits. A logical qubit consists of several physical qubits. In stabilizer codes, a popular class of quantum error correction schemes, a part of the system of physical qubits is measured repeatedly, without measuring (and collapsing by the Born rule) the state of the encoded logical qubit. These repetitive measurements are called syndrome measurements, and must be interpreted by a classical decoder in order to determine what errors occurred on the underlying physical system. The decoding of these space- and time-correlated syndromes is a highly non-trivial task, and efficient decoding algorithms are known only for a few stabilizer codes.

In our research we design and train decoders based on recurrent neural networks. The training is done using only experimentally accessible data. A key requirement for an efficient decoder is that it can decode an arbitrary and unspecified number of error correction cycles. To achieve this we use so-called long short-term memory cells [1]. These recurrent neural network building blocks have an internal memory. During training, the decoder learns how to update and utilize its internal memory, in order to detect errors on the logical qubit. The trained decoder is therefore a round based algorithm, rather than a rigid pattern recognition scheme. It can process the syndrome information in real-time, without having to wait for the quantum computation to be completed.

In our recent work [2] we have focused on one type of stabilizer code, the surface code, which is currently being implemented by several experimental groups [3-5]. We have trained and tested the neural network decoder on both a simple circuit model, and on a density matrix simulator with experimental parameters. In the presence of correlated bit-flip and phase-flip errors the neural network decoder outperforms the popular minimum-weight perfect matching decoder. However, our neural network decoder is not tailored to the specifics of the surface code, and should also be applicable to other stabilizer codes, such as the color code [6].
References

Quantum learning of coherent states

Gael Sentís, Mădălin Guţă, and Gerardo Adesso

EPJ Quantum Technology 2, 17 (2015), doi:10.1140/epjqt/s40507-015-0030-4

Quantum learning has strong links with quantum information theory and processing, as an area of strong foundational and technological interest, has recently raised great attention. Particularly, the use of programmable quantum processors has been investigated to address machine learning tasks such as pattern matching [4], binary classification [5–8], feedback-adaptive quantum measurements [9], learning of unitary transformations [10], & probably approximately correct learning [11], and unsupervised clustering [12]. Quantum learning algorithms provide not only performance improvements over some classical learning problems, but they naturally have also a wider range of applicability. Quantum learning has also strong links with quantum control theory [13], and is thus becoming an increasingly significant element of the theoretical and experimental quantum information processing toolbox.

In this paper, we investigate a quantum learning scheme for the task of discriminating between two coherent states. Coherent states stand out for their relevance in quantum optical communication theory [14, 15], quantum information processing implementations with light, atomic ensembles, and interfaces thereof [16, 17], and quantum optical process tomography [18]. Lasers are widely used in current telecommunication systems, and the transmission of information can be theoretically modelled in terms of bits encoded in the amplitude or phase modulation of a laser beam. The basic task of distinguishing two coherent states is thus of great importance, since lower chances of misidentification translate into higher transfer rates between the sender and the receiver.

The main goal of this paper is to explore the fundamental task of discriminating between two coherent states with minimum error, when the available information about their amplitudes is incomplete. The simplest instance of such problem is a partial knowledge situation: the discrimination between the (known) vacuum state, |0⟩, and some coherent state, |α⟩, where the value of α is not provided beforehand in the classical sense, but instead encoded in a number n of auxiliary modes in the state |α⟩⊗n. Such discrimination scheme can be cast as a learning protocol with two steps: a first training stage where the auxiliary modes (the training set) are measured to obtain an estimate of α, followed by a discrimination measurement based on this estimate. We then investigate whether this two-step learning procedure matches the performance of the most general quantum protocol, namely a global discrimination measurement that acts jointly over the auxiliary modes and the state to be identified.

In order to motivate further the problem investigated in this paper, let us define the specifics of the setting in the context of a quantum-enhanced readout of classically-stored information. Imagine a classical memory register modelled by an array of cells, where each cell contains a reflective medium with two possible reflectivities r₀ and r₁. To read the information stored in the register, one shines light into one of the cells and analyses its reflection. The task essentially consists in distinguishing the two possible states of the reflected signal, |0⟩ and |α⟩, assisted by n auxiliary modes sent directly by the transmitter.

FIG. 1. A quantum reading scheme that uses a coherent signal |α⟩, produced by a transmitter, to illuminate a cell of a register that stores a bit of information. A receiver extracts this bit by distinguishing between the two possible states of the reflected signal, |0⟩ and |α⟩, assisted by n auxiliary modes sent directly by the transmitter.
To aid in the discrimination of the signal, we alleviate the effects of the uncertainty in $\alpha$ by considering that $n$ auxiliary modes are produced by the transmitter in the global state $|\alpha\rangle^n_{\text{trans}}$ and sent directly to the receiver. The receiver then performs measurements over the signal and the auxiliary modes and outputs a binary result, corresponding with some probability to the bit stored in the irradiated cell.

We set ourselves to answer the following questions: (i) which is the optimal (unrestricted) measurement, in terms of the error probability, that the receiver can perform? and (ii) is a joint measurement, performed over the signal together with the auxiliary modes, necessary to achieve optimality? To accomplish the set task, we first obtain the optimal minimum-error probability considering collective measurements. Then, we contrast the result with that of the standard estimate-and-discriminate (E&D) strategy, consisting in first estimating $\alpha$ by measuring the auxiliary modes, and then using the acquired information to determine the signal state by a discrimination measurement tuned to distinguish the vacuum state $|0\rangle$ from a coherent state with the estimated amplitude. In order to compare the performance of the two strategies, we focus on the asymptotic limit of large $n$. The natural figure of merit is the excess risk, defined as the excess asymptotic average error per discrimination when $\alpha$ is perfectly known. We show that a collective measurement provides a lower excess risk than any Gaussian E&D strategy, and we conjecture (and provide strong evidence) that this is the case for all local strategies.


A fundamental problem of inference is that of the observation of a long (ideally infinite) stationary time series of events, generated by a hidden Markov chain. What can we say about the internal structure of the hidden Markov model, aka the latent variables? If the system generating the observations is classical, we are looking to reconstruct the “hidden” Markov chain from its “visible” image. Here, we are studying the case that the hidden system is quantum mechanical, giving rise to a special class of finitely correlated states, which we call “quantum hidden Markov models”; and even more generally, a generalised probabilistic theory (GPT). The latter case is entirely described in terms of the rank of the so-called Hankel matrix, and an associated canonical vector space with associated positive cone preserved under the hidden dynamics of the model. For the quantum case, we describe the structure of the possible GPTs via semidefinite representable (SDR) cones. It turns out that these GPTs are all finitely presented operator systems, i.e. induced subspaces of quotients of B(H) for a finite-dimensional Hilbert space H. Unlike operator systems, for which complete positivity can be very hard to decide, the SDR models come with a subset of the completely positive maps, which is itself an SDR cone, nicknamed ‘o Scarrafone (youtu.be/LCJKeAbHe8). My aim is to discuss what we know about the geometry of the GPT cones and this ugly/beautiful mapping cone, and how they relate to the operator system structure on the hand hand, and to general GPTs on the other.

[Based on joint work with Alex Monras, arXiv:1412.3634]
Towards Quantum-Assisted Artificial Intelligence
(Invited Talk)

Peter Wittek
Institute for Photonic Science
ICFO Barcelona, Spain

The recent progress in building scalable gate-model quantum computers and other quantum computing architectures has been remarkable. Equally remarkable is the progress in machine learning and AI, where deep learning has achieved supremacy in discriminative supervised learning scenarios. To move beyond that, we face major bottlenecks both in theory and in computational resources. For these reasons, machine learning and AI comprise an important applied field where quantum resources are expected to give a major boost. On a theoretical level, we can study potential speedups of quantum-enhanced protocols, which range between exponential and quadratic reduction in complexity. On a more pragmatic level, we may consider what can be implemented with current and near-future technology, particularly when it comes to computationally expensive algorithms such as probabilistic graphical models. Even a constant-time speedup can enable these models the same way graphics processing units enabled deep learning. In this talk, we review the trajectory of quantum machine learning and discuss recent results that shift the attention towards more general models in artificial intelligence.
Inductive supervised quantum learning


Alex Monrás¹, Gael Sentís²,∗ and Peter Wittek³,⁴

¹Física Teórica: Información y Fenómenos Quánticos, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain
²Naturwissenschaftlich-Technische Fakultät, Universität Siegen, 57068 Siegen, Germany
³ICFO-The Institute of Photonic Sciences, Castelldefels, E-08860 Spain
⁴University of Borås, Borås, S-50190 Sweden

I. INTRODUCTION

Real-world problems often demand optimizing over massive amounts of data. Machine learning algorithms are particularly well suited to deal with such daunting tasks: by mimicking a learning process, algorithms are able to handle vast amounts of data and infer approximate solutions to otherwise intractable optimization problems. Quantum machine learning [1–3], an emergent line of research that introduces quantum resources in learning algorithms, explores the interplay between the machine learning framework and quantum information processing from various angles. A strong emphasis has been put so far on quantum-enhanced classical machine-learning algorithms, where computational speed ups are sought by harnessing the power of quantum computing. Quantum mechanics, however, also alters the limitations on what is physically possible in a classical learning setup, thus potentially changing the structure of learning algorithms at a fundamental level and opening a door for increased performance and new phenomena. In particular, the handling of quantum data in a coherent manner has no classical counterpart, and typically leads to a performance enhancement over local incoherent approaches in many quantum information processing tasks [4–7]. On the other hand, when quantum data is considered, the no-cloning theorem readily hinders classical learning algorithms that need to process the same data several times. The integration of machine learning ideas into fully quantum setups that use both quantum data and quantum processors has been referred to as quantum learning. Establishing the ultimate limitations that quantum theory sets in learning algorithms is essential to assess, from a fundamental perspective, the potential advantages of quantum machine learning.

The study of supervised learning (arguably the most relevant class of algorithms) in a fully quantum scenario was pioneered in an early work of Sasaki and Carlini in 2002 [8]. In the context of quantum template matching, the authors showed that a semiclassical strategy, that consists in performing separate quantum measurements over the training data and the test data and coincides with a conventional (classical) supervised approach, strictly underperforms a coherent quantum measurement that acts on both types of data at once. Later works extensively studied the task of supervised classification of quantum states in various settings, and obtained differing results. In the case of classifying qubit states, the existence of a performance gap between a semiclassical strategy and a fully coherent one depends on the chosen figure of merit [9, 10]; if we consider coherent states, the gap exists [11]. All these results show that genuinely quantum features, such as performing coherent global operations in a multipartite system and the impossibility of cloning a quantum state, profoundly affect the structure and performance of quantum learning protocols, and illustrate the subtle nature of task-dependent knowledge acquisition in quantum scenarios.

II. MAIN RESULTS

In this talk I will first briefly overview previous results on supervised quantum learning from Refs. [8–11]. Then, I will develop a completely general framework for supervised quantum learning algorithms, following Ref. [Phys. Rev. Lett. 118, 190503 (2017)], and contrast its structure with its classical analog. I will first show for the classical case that a natural independence requirement among test data instances, i.e., that the learning algorithm be nonsignaling, induces the standard splitting of inductive learning algorithms into a training phase and a test phase.

∗Electronic address: gael.sentis@gmail.com
FIG. 1: Diagramatic representation of a generic quantum learning protocol $Q$ (grey box), which be approximated by $\tilde{Q}$ (black boxes inside the grey box). Both setups take training and test instances $\rho_A$ and $\rho^{\otimes n}_{XY}$ as inputs. We distinguish two agents in the diagrams: the performer of the learning protocol or “learner”, placed above the dashed horizontal line, and the “referee”, placed below. The learner sends the output registers $Y_1^{1:n}$ of the learning channel to the referee, who contrasts them with the registers $Y_1^{1:n}$ and evaluates the average risk of the channel $\bar{S}$. While the most general approach (a collective quantum channel $Q$) in principle acts globally on all its inputs, its approximation $\tilde{Q}$ comprises two separate phases: first (training phase) a measurement $M$ is performed on the training set $\rho_A$, and second (test phase) the classical information $g$ obtained from the measurement is distributed among all test instances, and corresponding quantum channels $\Phi_g$ are applied locally to each one of them.

I will then prove that the same splitting holds asymptotically in the quantum case, despite having access to coherent collective quantum operations (see Fig. 1). In other words, I will show that, in a fully quantum setting, the following three statements are equivalent in the asymptotic limit of large data sets in terms of performance:

1. a supervised learning algorithm learns a function which is applied to every test instance,
2. a supervised learning algorithm satisfies a nonsignaling criterion,
3. a supervised learning scenario splits into separate training and test phases.

The result is based on a de Finetti theorem for nonsignaling quantum channels. We use it to prove that the performance of any supervised quantum learning algorithm, under the restriction of being nonsignaling, approaches that of a protocol that first measures the training instances and then infers labels on the test instances, in the limit of many test instances. Our result reveals a natural analogy between classical and quantum supervised learning algorithms, justifying a similar treatment that we have been taken for granted. Ultimately, the result provides a solid basis to generalize key concepts in statistical learning theory, such as structural risk minimization [13], to quantum scenarios.

Adversarial Domain Adaptation for Identifying Quantum Phase Transitions

Patrick Huembeli, Alexandre Dauphin and Peter Wittek
ICFO, Barcelona, Spain

Abstract
The identification of phases of matter is a challenging task, especially in quantum mechanics, where the complexity of the ground state grows exponentially with the size of the system. The derivation of the phase diagram requires an efficient study of the parameter space. We address this problem with state-of-the-art deep learning techniques: adversarial domain adaptation. We derive the phase diagram of the whole parameter space starting from fixed and known subspace in an unsupervised way. More specifically, the input is both labelled data instances and the unlabelled data instances. The first type is a system that is analytically or numerically easy to treat, and hence we know its phase diagram: the phases are the labels in machine learning parlour. The second type is the physical system for which we would like derive the phase diagram. Adversarial domain adaptation uses both types of data to create invariant feature extracting layers in a deep learning architecture. Once these layers are trained, we can attach an unsupervised learner to the network to find phase transitions. We show the success of this technique by applying it on several paradigmatic models: the Ising model with different temperatures, the Bose-Hubbard model, and the SSH model with disorder. The method is promising as it can be applied directly on the raw ground state, without any manual feature engineering, and can be used on a parameter space of arbitrary dimensions. Future directions would be to classify physical systems where the phases boundaries are complex such as the many body localisation problem or the Bose glass phase.
Quantum Change Point Identification

John Calsamiglia, Emilio Bagan, Ramon Muñoz-Tapia and Gael Sentis

GIQ- Universität Autonoma de Barcelona, Spain
Universität Siegen, Germany

Abstract

Without interpretation it is hard for humans to make sense and make predictions from raw data. Machine learning provides algorithmic tools that allow computers to make sense out of (big) data. However, data itself suffers changes and it is hence of paramount interest to develop schemes to learn pattern changes and to identify the change points.

Moving away from a standard machine learning framework of large and unstructured data sets, change point detection is a pivotal task also in traditional statistical inference disciplines, where data has typically smaller dimensionality and some prior knowledge is usually assumed. The field of change point detection aims at identifying the moment when a sequence of observed data changes its underlying probability distribution. The problem has widespread applications, including manufacturing (quality control), intrusion detection, stock market variations, protein folding, and landscape changes.

We study the quantum extension of this problem where a source emits quantum particles in a default state, until a point where a mutation occurs that causes the source to produce a different state. The problem is then to find out where the change occurred. We determine the optimal strategy for the correct identification of the change point with minimum error probability, as well as for the unambiguous identification, where the change point is identified with certainty—at the price of getting an inconclusive answer with some non-vanishing probability. For the latter we give the exact expression for the optimal success probability valid for all (not necessarily asymptotically large) number of samples. This is remarkable because there are very few known instances of exactly solvable problems of unambiguous multi-hypothesis discrimination. We also discuss local protocols and compare them with the optimal procedures.

The contents of this talk are available at quant-ph arXiv:1707.07769 [to appear in PRL] and arXiv:1605.01916 [Phys. Rev. Lett. 117, 150502 (2016)]. If time permits I will also discuss ongoing work on the QUSUM algorithm: a the quantum extension of the classical CUSUM algorithm that solves the quickest change-point detection problem. In this setting one wishes to minimize the response time (time delay between actual change-point and its detection) given a fixed false-alarm rate.
A Quantum-inspired version of the Nearest Mean Classifier

Giuseppe Sergioli, Ranjith Venkatrama, Roberto Giuntini

University of Cagliari

We introduce a new framework for describing pattern recognition tasks by means of the mathematical language of density matrices. In recent years, many efforts to apply the quantum formalism to non-microscopic contexts have been made and, in this direction, important contributions in the areas of pattern recognition and image understanding have been provided [1,2,5,6,7]. Even if these results seem to suggest some possible computational advantages of an approach of this sort, an extensive and universally recognized treatment of the topic is still missing [6,4,3].

The natural motivations which have led to use quantum states for the purpose of representing patterns are i) the possibility to exploit quantum algorithms to boost the computational intensive parts of the classification process, ii) the possibility of using quantum-inspired algorithms for a more effectively solving of classical problems. In our work, firstly we provide a one-to-one correspondence between patterns, expressed as n-dimensional feature vectors (according to the standard pattern recognition approach), and pure density operators (i.e. points in the n-dimensional Bloch hypersphere) called "density patterns". By using this correspondence, we give a representation of the well-known Nearest Mean classifier (NMC) in terms of quantum objects by defining an "ad hoc" Normalized Trace Distance (which coincides with the Euclidean distance between patterns in the real space). Consequently, we introduce a quantum version of the discriminant function by means of Pauli components, represented as a plane which intersects the Bloch sphere. In this way, what we obtain is a quantum version of the NMC. This first result suggests future potential developments, which consist in finding a quantum algorithm able to implement the normalized trace distance between density patterns with a consequent significative reduction of the computational complexity of the process. But the main result that we show is based in the introduction of a purely quantum classifier (QC), which has not any kind of classical counterpart, through a new definition of "quantum centroid". The convenience of using this quantum centroid lies in the fact that it seems to be more informative than the classical one because it takes into account also information about the distribution of the patterns. As a consequence, the main implementative result consists in showing how this quantum classifier performs a significative reduction of the error and an improvement of the accuracy and precision of the algorithm with respect to the NMC (and also to other commonly used classifiers) on a classical computer (i.e. without involving quantum algorithms).

The behaviors of QC and NMC on different datasets are shown and the values of the error are compared. We will show how the QC exhibits a relevant improvement in the reduction of the classification error for all the considered datasets. Finally, we observe as, unlike the NMC, the QC is not invariant under the rescaling of the patterns; i.e., the error changes by rescaling the features of the patterns of a given dataset. We show how it can be used as a ulterior advantage that allows to obtain a further improvement in the reduction of the classification error. Some of these preliminary results have been recently published [8,9,10,11].
References:


Quantum Machine Learning with Small-Scale Devices  
(Invited Talk)  

Francesco Petruccione  
University of KwaZulu-Natal, South Africa  

Lately, much attention has been given to quantum algorithms that solve pattern recognition tasks in machine learning. Many of these quantum machine learning algorithms try to implement classical models on large-scale universal quantum computers that have access to non-trivial subroutines such as Hamiltonian simulation, amplitude amplification and phase estimation. We approach the problem from the opposite direction and analyse a distance-based classifier that is realised by a simple quantum interference circuit. After state preparation, the circuit only consists of a Hadamard gate as well as two single-qubit measurements and can be implemented with small-scale setups available today. We demonstrate this using the IBM Quantum Experience and analyse the classifier with numerical simulations.
Demonstration of Envariance and Parity Learning on the IBM 16 Qubit Processor

Davide Ferrari\textsuperscript{1} and Michele Amoretti\textsuperscript{1,2}

\textsuperscript{1}: Department of Engineering and Architecture - University of Parma, Italy
\textsuperscript{2}: Quantum Information Science @ University of Parma, Italy

http://www.qis.unipr.it

Contacts: davide.ferrari8@studenti.unipr.it, michele.amoretti@unipr.it

Extended Abstract

Since 2016, IBM offers hands-on, cloud-based access to its experimental quantum computing platform, denoted as IBM Quantum Experience [1]. Such a platform comprises a 5 qubit device, denoted as IBM Q5, and a 16 qubit device, named IBM Q16. Both devices are based on transmon qubits [2], i.e., superconducting qubits which are insensitive to charge noise. IBM Quantum Experience is calibrated daily and the decoherence times of its qubits are about 100µsec. A Web-based visual tool provides a convenient way to compose quantum circuits for IBM Q5 and run either simulated or real experiments. Alternatively, circuits can be designed by means of the QASM language and experiments can be executed by means of the QISKit Python SDK. Actually, IBM Q16 can be accessed in this way only.

The research community has welcomed the IBM Quantum Experience as an experimental platform for testing quantum algorithms. Current official numbers are: 50000 users, 500000 experiments and more than 25 papers [3].

Deffner [4] recently reported a simple and easily reproducible demonstration of entanglement assisted invariance, also known as envariance [5], using IBM Q5. In this work, we present our experience in demonstrating envariance just as Deffner did, but with more qubits, using IBM Q16. In particular, we illustrate the non-trivial strategy we have designed for automatically composing circuits that produce GHZ states, taking into account the topological constraints of IBM Q16.

Machine learning techniques are powerful tools for finding patterns in data. The field of quantum machine learning looks for faster machine learning algorithms, based on quantum computing principles [6]. Its cornerstones are the HHL algorithm [7] for (approximately) solving systems of linear equations and the “learning from quantum examples” approach [8, 9, 10, 11], where each example is a coherent quantum state.

As a matter of fact, learning a class of Boolean functions by making queries to a uniform quantum example oracle can be effectively implemented on IBM Quantum Experience. Cross \textit{et al.} [12] proved that parity learning can be performed with superpolynomial quantum computational speedup only in the presence of noise. The experimental demonstration on IBM Q5 was recently presented by Ristè \textit{et al.} [13]. In this work, we show how we performed similar experiments on IBM Q16, also in this case using a non-trivial strategy to automatically compose the circuits.

The source code we have implemented is available on GitHub [14].
Acknowledgements

The authors acknowledge the use of IBM Quantum Experience for this work. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum Experience team. This work has been supported by the University of Parma Research Fund - FIL 2016 - Project “NEXTALGO: Efficient Algorithms for Next-Generation Distributed Systems”.

References

I. INTRODUCTION

Recent advances in the development of quantum information processor (QIP) hardware, such as new multi-qubit devices at IBM [1] and Rigetti Quantum Computing [2], have led researchers to consider how large-scale QIPs will be characterized. Although current techniques (such as tomography and randomized benchmarking) work well for current QIPs, new techniques will need to be devised as the number of qubits grows. Eventually full-scale tomography (which provides maximal information about noise) will no longer be feasible, as the number of parameters to be estimated in a density or process matrix grows exponentially with the number of qubits. Further, although randomized benchmarking is easy, the information it provides about noise properties is limited. The discrepancies between speed of learning and amount of information learned motivate the question “What are characterization techniques that learn useful information about noise while being fast to compute?”.

Here, I show that principal component analysis and support vector machines occupy that “happy middle”, allowing for the quick extraction of some useful information about noise. This is because noise can affect the outcome probabilities of quantum circuits. For example, suppose that a single-qubit QIP has a noise process such that before the POVM \( \{ \ket{0}, \ket{1} \} \) is measured, an X-error (\( \pi \) rotation) occurs. Suppose further that the QIP performs a circuit such that the state of the qubit prior to measurement is \( \ket{0} \). In the absence of this noise, \( \Pr(\ket{0}) = 1 \), while when the noise is present, \( \Pr(\ket{0}) \) drops to 0.

Of course, there are many noise processes that affect QIPs beyond the one just mentioned, and ill-chosen circuits may not detect them all. Gate set tomography (GST) [4–7] is a tomographic protocol that has been designed to address this issue, by finding circuits that amplify all noise. The full details of this design procedure are beyond the scope of this abstract; for our purposes, it suffices to know two things. First, GST assumes the QIP may be described by a gate set \( \mathcal{G} = \{ \rho, \{ G_j \}, \{ E, I - E \} \} \), where \( \rho \) is the state produced by the initialization procedure of the device, \( \{ G_j \} \) is a set of quantum gates the device may perform (e.g., rotations about the \( X, Y, Z \) axes), and \( \{ E, I - E \} \) is a two-outcome POVM. Second, the circuits selected by GST, denoted \( \{ c_j \} \), comprise a state preparation, the application of a composition of subsets of the gates \( \{ G_j \} \), and then a measurement. These circuits have the property that their total length (the number of primitive gates performed) is exponentially spaced, typically increasing as \( 1, 2, 4, \ldots \log_2(L) \), for some cutoff length \( L \) [3].

A GST dataset is a list \( \{ (c_1, f_1), (c_2, f_2), \ldots \} \), where \( c_j \) denotes the circuit, and \( f_j \) is the frequency that outcome \( E \) was observed. GST processes this dataset to return an estimate, \( \hat{\mathcal{G}} \), of the gate set whose error scales as \( O(1/L) \). Unfortunately, computing that estimate is time-intensive, taking around 24 hours for a 2-qubit QIP. However, because the circuits GST chooses amply all noise, all the information about noise is encoded in that dataset.

II. DATASETS AS FEATURE VECTORS

The observed frequencies from a GST dataset form a vector \( f = [f_1, f_2, \ldots] \) in a feature space whose features are the circuits themselves. This feature space is a subset of \( \mathbb{R}^d \) [8], where \( d \) (the number of circuits) is determined by \( L \). Because these feature vectors contain all the information about the noise, it stands to reason that principled machine learning tools can extract a subset of that information.

To explore this possibility, I created 114 synthetic single-qubit GST gate sets using the open source pyGSTi software [10]. They have one of three noise sources (depolarization, coherent error, or no noise at all), noise strengths in \( [0.005, \ldots, 0.045] \), and a value of \( L \) in \( [1, \ldots, 64] \). For each fixed noise type, noise strength, and value of \( L \), 50 unique feature vectors were generated, by using binomial sampling of the true probabilities of each circuit \( c_j \), with the number of trials \( n = 100 \).
III. PRINCIPAL COMPONENT ANALYSIS OF FEATURE VECTORS

The dimension of the feature space grows with \( L \), because GST adds more and more circuits. By using principal component analysis (PCA), it is possible to find a lower-dimensional space which provides a useful approximation to the data [11–13]. This approximation allows us to better visualize how the circuits chosen by GST affect properties of the feature vectors, such as their distinguishability.

PCA creates such an approximation by determining the directions of maximal variance within the data. Given a collection of \( N \) feature vectors \( \mathbf{f} \in \mathbb{R}^d \), PCA returns a list of \( K \leq \min(d, N) \) orthonormal vectors \( \{\mathbf{e}_j\} \), known as the principal components. The original data can be approximated in a lower-dimensional space as \( \mathbf{f} \rightarrow \mathbf{f}' = \sum_{j=1}^{K} \mathbf{e}_j (\mathbf{e}_j^T \mathbf{f}) \). Thus, in the basis defined by \( \{\mathbf{e}_j\} \), \( \mathbf{f} \) now has dimension \( K \).

It itself is a hyperparameter of the learning process, and can be determined empirically [9]. For ease of visualization, however, I keep only the first two components \( (K = 2) \). Figure 1 plots the projections \( \mathbf{f}' \), which show interesting behavior. They cluster [14], and as \( L \) increases, they become further apart. Examining the clusters by the noise type, different noise types tend to be projected to different parts of this space. Thus, using PCA – even with \( K = 2 \) – allows us to create an approximation of GST feature vectors which is (a) simpler than the original feature vectors, (b) easily comprehended, and (c) useful for classification purposes (the subject of the next section).

IV. USING SUPPORT VECTOR MACHINES TO CLASSIFY NOISE

Support vector machines (SVMs) are a well-established tool for classification tasks [17]. In its simplest form, an SVM finds the optimal (“maximal margin”) hyperplane(s) for the data. Generalizations are possible (e.g., via kernel methods [18]), but here I focus on the simple case of a soft-margin, linear SVM. Given \( N \) training vectors \( \mathbf{f} \) with two possible labels (without loss of generality, \( \{\pm 1\} \)), the SVM learns a weight vector \( \mathbf{w} \) and a bias \( b \), where \( \mathbf{w} \) and \( b \) are chosen to minimize the inaccuracy of the SVM, subject to regularization of \( ||\mathbf{w}||^2 \). A new vector \( \mathbf{z} \) is then classified as \( \mathbf{z} \rightarrow \text{sgn}(\langle \mathbf{w} \cdot \mathbf{z} + b \rangle) \).

Because the data used here has 3 possible labels, the classification scheme above must be modified. I chose to use a “one-versus-one” scheme, so each pair of labels is learned by the SVM.

Figure 2 compares the cross-validated performance of the SVM on the entire feature space, as well as on the approximations returned by PCA. The SVM attains a slightly higher accuracy without PCA, which makes sense, as PCA can throw away information about the feature vectors. However, the decrease in accuracy is modest when using PCA, and the approximations it creates can be much more easily visualized and understood.

V. CONCLUSIONS/FUTURE WORK

In this exploratory work, I showed that machine learning tools can be used to learn about properties of noise affecting QIPs. By using the outcomes of carefully chosen circuits, and interpreting those datasets as feature vectors, PCA can be used to reduce the dimensionality of the feature space, as well as provide a useful understanding of why those circuits enable learning of noise properties. Using SVMs, it is possible to classify (a highly restricted set of) single-qubit noise sources with good accuracy.

It would be productive to address whether support vector regressors (SVRs) can learn the strength of such noises as well. Further, a fruitful avenue of inquiry would be to tackle the issue of circuit design to further tailor the circuits to the kind of noise being learned.

Of course, no QIP is ever affected by only one noise source, and a Markovian one at that. Investigating the feasibility of learning multiple, concurrent noise sources, as well as non-Markovian noise, is also necessary.
VI. ACKNOWLEDGEMENTS

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525.

[3] In the example given above, any rotation which maps the state onto the equatorial plane of the Bloch sphere manifests itself in the data the same way. Thus, the circuit described can only verify that such a rotation has occurred, but it cannot help us learn around which axis the rotation took place.
[8] For reference, at $L = 1$, GST selects 92 sequences, and at $L = 64$, it selects 1969.
[9] This space is actually a subset of the unit hypercube in $\mathbb{R}^d$.
[14] For example, by evaluating the performance of the mean-squared error as a function of $K$.
[15] The centers of the clusters are determined by the noise strength, and their spread by the statistical sampling error.
Quantum enhanced neural network architecture evaluation

Adenilton José da Silva*

In this work, we describe a quantum algorithm that uses quantum parallelism to evaluate neural networks architectures with different number of neurons in hidden layer. Several works propose techniques to search for near-optimal neural networks architectures [1, 2, 3, 4]. One limitation of these works is the impossibility to evaluate a neural network architecture without many random weights initialization [2]. The quantum algorithm proposed in this section performs a neural network architecture evaluation that does not depends on a weights initialization because quantum parallelism is used to initialize the neural network weights with all possible weights.

Classically the main idea of this algorithm is to train all possible neural networks with a given architecture and create a binary vector performance for each weight vector \( w_j \). The performance has the same size of a validation dataset \( \{ (x_k, d_k) \} \) and the position \( k \) of the vector is equal to 1 if the network with weights \( w_j \) correctly classify \( x_k \). The algorithm compares the performance of each neural network with the performance representing an accuracy of 100% as described in Eq. (1), where \( t_s \) is the number of patterns in the validation set, \( |W| \) is the number of weights in superposition and \( d_H (\cdot) \) calculates the hamming distance.

\[
\sum_{k=1}^{|W|} \frac{1}{|W|} \cos^2 \left( \frac{\pi}{2t_s} \cdot d_H ([1]_{t_s}, \langle \text{performance}_k \rangle) \right)
\]

In a classical computer, this strategy cannot be accomplished in polynomial time. We show that in a quantum computer this algorithm can be executed in polynomial time and the result is related to the neural network architecture capacity to learn a dataset.

The quantum version of the proposed algorithm consists in create a superposition of all neural networks. Train the neural networks in the superposition. Create a performance vector for each neural network in the superposition and use the recovering algorithm of the quantum associative memory [5] using the performance quantum register as the quantum memory and with input \( |1\rangle_{t_s} \) to evaluate the neural network architecture. This idea is precisely described in Algorithm 1.

Line 4 trains the neural networks in the superposition. The learning algorithm of a classical neural network can be viewed as a binary function that sends the input \( x(t), \) desired output \( d(t) \) and weights \( w(t) \) in iteration \( t \) to the new weight vector \( w(t + 1) = f (w(t), x(t), d(t)) \). Any binary function can be simulated by a quantum operator \( U_f \) then the learning procedure can be simulated in a quantum computer and it is possible to train several artificial neural networks using quantum parallelism.

In line 13 the recovering algorithm of the probabilistic quantum memory is applied to the performance quantum register as memory and input \( |1\rangle_{t_s} \). The input \( |1\rangle_{t_s} \) represents a performance of 100%. Line 14 returns the control bit \( |c\rangle \) of the probabilistic quantum memory. A measurement of \( |c\rangle \) will return 0 with high probability if the neural network architecture is capable of learning the dataset.

The neural network simulation and algorithm output were simulated on a classical computer. To perform an evaluation of Algorithm 2 we follow its classical description and performs modifications of the number of neural networks in the quantum parallelism. This modification was necessary due to the high cost of train neural networks with all possible weights initialization. Instead of using all possible weights, we randomly select 1000 neural network weight vectors for a single layer neural network with \( k \) hidden neurons.

*Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Brazil
Algorithm 1: Evaluate architecture

1. Initialize all weights qubits with $H|0\rangle$
2. Divide the dataset $T$ in a train set and a validation dataset
3. Initialize quantum register $|\text{performance}\rangle$ with the quantum register $|0\rangle_n$
4. Train the neural networks in superposition with the train set
5. for each pattern $p_j$ in and desired output $d_j$ in the validation set do
   6. Initialize the quantum registers $p, o, d$ with the quantum state $|p, 0, d\rangle$
   7. Calculate $N|p_k\rangle$ to calculate network output in quantum register $|o\rangle$
   8. if $|o\rangle = |d\rangle$ then
      9. Set $|\text{performance}\rangle_j$ to 1
   10. end
   11. Calculate $N^{-1}$ to restore $|o\rangle$
12. end
13. Apply the quantum associative recovering algorithm with input $|1\rangle_n$ and memory $|\text{performance}\rangle$
14. Return the control qubit $|c\rangle$ of the quantum associative memory

Figure 1: Artificial neural network performance versus output of QNNAE algorithm $P(|c\rangle) = |0\rangle$ in datasets (a) cancer, (b) card

In future works, we intend to create a parametric quantum associative memory to improve the result described in Figure 1. We also intend to use a parametric quantum associative memory to perform pattern recognition [6].

References
Learning and Controlling a Spin Environment through Quantum Measurements

Durga Bhaktavatsala Rao Dasari, Stefan Jesenski, Johannes Greiner, Florian Kaiser and Joerg Wrachtrup
University of Stuttgart, Germany

Extended Abstract
Quantum systems are inevitably coupled to their surrounding environment and are never completely isolated. Learning such systems through quantum measurements not only affects the system of interest but also its surroundings i.e, environment influences the measurement result and in turn the measurement influences the environment. This kind of back-to-back action adds new feature in learning physical systems through quantum measurements. We explore this phenomena both theoretically and experimentally. We have used solid state defect centers in diamond as our model system for theory calculations [1, 2] and experiments. Our studies indicate that there exists an optimal state of the environment which influences (optimizes) the behavior of the system in a desired fashion [2]. Such states are shown to be ground states (minimum eigenvalue eigenstates) of a nonunitary operator which are otherwise computationally (NP) hard to obtain by numerical diagonalization methods. While this mimics the ground state stabilization of the system through adiabatic (unitary) evolution, using the proposed scheme we can generate ground states of various nonunitary operators only by changing the measurement basis in which the system is measured instead of varying the Hamiltonian parameters. As we will show this allows for robust error detection in the final state. The (intelligent) states of the environment which determine the behavior of the quantum system can be mapped to the ground states of strongly interacting systems opening new ways for engineering the quantum state of a macroscopic ensembles by projectively measuring (learning) its coupled partner.

References
Extractable Work-Based Criteria and Geometric Quantum Speed Limits for Quantum States Discrimination

Luca Mancino

Università degli Studi Roma Tre, Italy

Extended Abstract

Machine Learning (ML) is one of the most promising research avenues in computer science nowadays. It deals with establishing dynamical algorithms with which computers can recognize patterns learnt from data, without simply obeying fixed simple sequences of instructions. A great effort has been given to understand how to improve ML algorithms as the global data size is going to make them highly non-optimal. Using some of the central tools of quantum information processing, Quantum Machine Learning (QML) might revolutionize future ways of data processing, keeping advantage from purely quantum system properties, such as coherence and entanglement, quantum control and efficient representation of quantum states. Since Landauer’s principle introduces a physical link between the information processing and thermodynamic costs, entering the energetic balance of any physical process, thermodynamics-based protocols can be very interesting testbeds to optimise QML algorithms, such as state discrimination and pattern recognition. Within such a framework, Quantum Thermodynamics (QT) aids in witnessing entanglement, as well as in controlling transient dynamics, and evaluating irreversible entropy production during quantum systems evolution, providing on building blocks of quantum information processors. Quantum Speed Limits (QSLs), on the other hand, permit to efficiently represent quantum states, dictating how well quantum states can be discriminated by introducing metric-dependent dynamical paths and geodesic lengths on Riemannian manifolds. More in detail, here we want to show how thermodynamic protocols based on Maxwell’s demons can represent viable tests for the discrimination ability between separable and entangled states, also providing for the possibility to detect the entanglement type and robustness in multipartite schemes. Beyond entanglement discrimination, QT allows to distinguish quantum states undergoing dynamical evolution, as that dictated by a Lindblad-Davies master equation, exploring non-equilibrium quantum states. In such a scheme, QSLs permit both to identify the system path evolution on the
state space, permitting to discriminate states by evaluating distances. Our recent exertions aim at exploring the link between thermodynamic properties of a single-qubit undergoing thermalization and the Wigner-Yanase geodesic lengths on the Riemannian manifold, demonstrating how QSLs can assure lower bounds on the irreversible entropy production. Similar upper bounds can be introduced for the maximal rate of quantum learning, representing an important application of the quantum speed limit time. These theoretical predictions have been associated to quantum photonic laboratory experiments to show the robustness of such QT- and QSLs-inspired approaches for quantum information processing, and to contribute to technological advances.
Programming Quantum Computers: functional languages and new perspectives

Luca Paolini¹, Margherita Zorzi²

¹Department of Computer Science, University of Torino (Italy)
²Department of Computer Science, University of Verona (Italy)

September 16, 2017

Abstract

qPCF is a paradigmatic programming language that provides flexible facilities to manipulate quantum circuits. qPCF follows the tradition of “quantum data & classical control” languages, inspired to the QRAM model, where no intermediate quantum state is permanently stored and the emphasis is moved from states to circuits. qPCF enjoys properties such as Preservation and Progress Theorems and retains PCF’s expressive power: in fact, a qPCF terms can encode an (infinite) quantum circuit family.

1 Introduction

Quantum computers are a long term but tangible reality. Even if physicists and engineers have to continuously face tricky problems in the realization of quantum devices, the advance of these innovative technologies presents a noticeable speedup. In the last fifteen years, the definition and the development of quantum programming languages catalyzed the attention of a part of the computer science research community. The so called quantum data & classical control (qd&cc) approach separate the pure unitary quantum computational steps from the classical ones: a classical program (ideally in execution on a classical machine) computes some “directives”, sent to a hypothetical quantum device which applies them to quantum data. This idea is inspired to an architectural model called Quantum Random Access Machine (QRAM). In an hypothetical evolution of quantum technology, the qd&cc approach is definitely the more realistic one. We propose a new contribution in the research on the languages for qd&cc paradigm, in the line of project as Quipper or the recently developed quantum circuit definition language QWIRE. We formalize a flexible quantum language, called qPCF based, on a version of the QRAM model, restricted to total measurements. qPCF is a simple extension of PCF, and make quantum programming very easy.

2 qPCF

Ideally, qPCF computes a finite circuit description which is offloaded to a quantum co-processor for the execution. It extends PCF with a new kind of classical data, viz. quantum circuits. We summarise here its main features:

*Absence of explicit linear constraints*: the management of linear resources is radically different from the ones proposed in languages inspired to Linear Logic; so, we do not use linear/exponential modalities.

*Use of dependent types*: we decouple the classic control from the quantum computation by adopting a simplified form of dependent types that provide a sort of linear interface.

*Emphasis of the Standardization Theorem*: the Standardization Theorem decomposes computations in different phases, according to the quantum circuit construction by classical instructions and the successive, independent, evaluation phase claiming quantum operations.

*Unique measurement at the end of the computation*: following the “principle of deferred measurement”, we “embed” both quantum evaluation and a (von Neumann) total measurement into an operator called dMeas and completely relieving the programmer of the tricky management of quantum data.

*Turing Completeness*: A qPCF term is able to represent an infinite class of circuits. When fed with a numeral, the evaluation returns, among the infinite family, exactly the description of the circuit for the given size.

**Properties.** qPCF enjoys some good properties such as *Preservation* and *Progress* Theorems.

**Example.** qPCF permits to easily encode (parametrically w.r.t. the input dimension) some well-known quantum algorithm. The following qPCF terms encodes different levels of the measurement-free parametric circuit for the Deutsch-Jozsa problem. Let $H : \text{circ}(0)$ and $I : \text{circ}(0)$ be the (unary) Hadamard and Identity gates respectively (so the index is 0). Suppose $M^{\ast_{f}} : \text{circ}(0)$ is given for some $n$ such that $M^{\ast_{f}} \downarrow U_{f}$ where $U_{f} : \text{circ}(n)$ is the qPCF-circuit that represents the function $f$. Observe that $\lambda x^{\text{circ}}. \text{iter} x H H : \Pi^{\text{circ}}. \text{circ}(x)$ clearly generates $x + 1$ parallel copies of unary Hadamard gates $H$, and $\lambda x^{\text{circ}}. \text{iter} x I H : \Pi^{\text{circ}}. \text{circ}(x)$ concatenates in parallel $x$ copies of unary Hadamard gates $H$ and one copy of the unary identity gate $I$. Thus the generator term of the parametric measurement-free Deutsch-Jozsa circuit, here dubbed $DJ^{-}$ can be defined as $DJ^{-} = \lambda x^{\text{circ}}. \lambda y^{\text{circ}}. ((\text{iter} x H H) \% y) \% (\text{iter} x I H) : \sigma$ where $\sigma = \Pi^{\text{circ}}. \text{circ}(x) \rightarrow \text{circ}(x)$. We finally evaluate $DJ^{-}$ by means of $\text{dMeas}$, providing the encoding $M^{\ast_{f}} \circ \text{circ}(0)$ of the black-box function $f$ having arity $n + 1$. Let us assume that the term $\text{dMeas}(\text{setto}_{p} DJ^{-} \# M^{\ast_{f}})$ evaluated by means of $\psi_{p}$, yields the numeral $\overline{n}$ the rightmost $\overline{n}$ digit of the binary representation of $\overline{n}$ are the result.

3 Discussions and conclusions

The presented research is open to several short time investigations. First, we aim to deepen qPCF flexibility, e.g. studying *specialization* of qPCF. We like to note that gates included into the syntax can range on different interesting sets e.g. reversible circuits. Even if the use of total measurement does not represent a theoretical limitation, a partial measurement operator can represent a useful programming tool. Therefore, another interesting task will be to integrate in qPCF the possibility to perform partial measures of computation results.
Modelling Multipartite Entanglement in Quantum Protocols using Evolving Entangled Hypergraphs

Linda Anticoli and Masoud Gharahi Ghahi

1 Dept. of Mathematics, Computer Science and Physics, University of Udine, Italy
2 masoud.gharahi@gmail.com

Modelling Multipartite Entanglement in Quantum Protocols using Evolving Entangled Hypergraphs

Linda Anticoli1 and Masoud Gharahi Ghahi2

1 Dept. of Mathematics, Computer Science and Physics, University of Udine, Italy
2 masoud.gharahi@gmail.com

Multi-partite Entanglement Classification via Hypergraphs: The notion of entanglement is a core concept in Quantum Mechanics. It is a feature displayed by a certain class of quantum states and also a powerful resource for quantum information and computation, with applications ranging from quantum teleportation to more complex tasks such as quantum cryptography and quantum key distribution protocols (QKD). Thus, entanglement as a resource needs to be classified since each class does the same task in quantum information processing. Its classification, however, is not trivial.

Quantum entanglement in bipartite systems of pure states is (almost) completely understood because it can be characterized using the Schmidt decomposition. Thus, by means of local unitary transformation only we can write any pure state in a bipartite system in a canonical form with non-local properties contained in the positive Schmidt coefficients, which tell us whether the state is separable or entangled.

Quantum entanglement in multipartite systems of pure states is not easy to classify as in the bipartite case. Indeed, in multipartite entanglement, which refers to correlations between more than two particles, it is not enough to know whether the subsystems are entangled or not, but also how they are entangled. There are, indeed, different ways in which a pure state \( |\psi\rangle \in \mathcal{H}^{(1)} \times \cdots \times \mathcal{H}^{(N)} \) can be entangled. In tripartite systems, for example, we can have separable states, biseparable states and two kinds of (locally inequivalent) tripartite entangled states, i.e., GHZ states and \( W \) states, which cannot be transformed in each other by means of stochastic local operations and classical communication (SLOCC) [1]. In the case of multipartite entanglement exists a generalized Schmidt decomposition (GSD) [2] which will be used by the classification approach considered in this work. GSD allows us to put any state in a canonical form, which is then used as a starting point for other algorithmic steps, involving entanglement measures such as concurrence, tangle and global entanglement. In this work we extend the idea presented in [3] which was in turn based on the concept of entangled graphs [4].

An entangled graph is a structure in which each vertex represents a qubit and each edge between two qubits represents bipartite entanglement. In this work we introduce the notion of entangled hypergraphs (EH), a generalization of entangled graphs suitable to be applied in the context of multipartite entanglement; in an EH each vertex \( v_i \) represents a qubit, and each hyperedge \( e_i \) connecting a set of vertices represents multipartite entanglement. We will propose an algorithmic framework based on GSD and the entanglement measures mentioned before to classify multipartite entanglement of pure states using EH. Moreover, by starting from the equivalence classes modeled with EH, we can use an algebraic framework to write the pure state associated to each class.

We have found that some EHs are forbidden, since we cannot write a pure state describing them. In the three qubits case, for example, the star shape hypergraph without 3-edge is forbidden, as in the figure on the left in this page.

This method works as expected in the tripartite case; some problems arise in this classification when we generalize it to the \( \geq 4 \)-partite case. Anyways, we argue that this method can be a good starting point for the investigation of other multipartite classification techniques, since it has a finitary approach, which is suitable for further modeling and verification tasks.
Evolving EH for Modeling Quantum Protocols: The core part of our work is devoted to use the classification introduced as a starting point to model quantum protocols which use entanglement as a resource. We propose the concept of Evolving Entangled Hypergraph (EEH), a structure inspired by both EH and hypergraph states, taking into account not only correlations (entanglement) but also interactions (evolution) of the particles under consideration. Interactions in this case are represented as a dynamical process, i.e., the evolution of the quantum system in successive, discrete time steps. This approach turns out to be useful in particular when we try to model quantum protocols in which entanglement is an important resource to be preserved, e.g., QKD protocols and teleportation.

We give the definition of a EEH as a causal multilayer hypergraph in which each layer $L_i$ represents the EH at time $t_i$ (i.e., one of the allowed classes) and the hyperedges from a layer to the following one are labeled with the CPTP (completely positive and trace preserving) linear map acting on the states of EH. An example, using the tripartite scenario, of a possible evolving hypergraph can be seen in the figure in this page, where $\rho$ refers to a given state of three particles and the transformation $E$ and $F$ are CPTP maps.

In this example entanglement between $v_1$ and $v_2$ is preserved at time $t_1$, then the channel $F$ change something causing decoherence.

The EEH model allows us to track also the interactions between qubits of a multipartite system, and thus to verify if the entanglement properties are preserved by the time evolution. Since we want to model real quantum systems, we assume that the evolution may not be unitary, and that some coupling with the environment may cause noise and decoherence, while entanglement is supposed to be preserved for the success of the quantum protocol under consideration. Moreover, we can also use forbidden EHs to identify whether the protocol is valid or not.

This model is an abstract structure taking in input a class $L_i$ from a finite set $L$ of equivalence classes, evolving according some CPTP $E$ and giving in output another class from the same set, i.e., $\{L_i\} \xrightarrow{E} \{L_j\}$ with $L_i, L_j \in L$. In this way, the only constraint we impose is that the method used to classify multipartite entanglement is finitary. We suppose that the structures proposed can be used in model-checking as models which can be formally verified by using, e.g., Spatial-Temporal logics as [5], since we will be able to both define spatial locations of qubits (i.e., the vertices of the EH can be interpreted as qubit positions) and the temporal behavior of the system. In the future we would try to investigate whether machine learning techniques can be used to assess if the entanglement property is preserved or not within the EEH.

The advantages in using the entanglement classification with EH is that it is a finitary method and that it is not required to compute the mathematical representation of operators. The hypergraph is computed by an algorithmic procedure, for which we are planning to build a tool to automate the process. To extract information about the entanglement from the system, modeled as an EEH, we can work on an high level structure. Then, by having an explicit classification declaring which are the allowed and forbidden configurations, we can verify whether the structure matches the requirements.

Further investigations should be done in different directions, in order to provide a reliable classification of multipartite entanglement for systems with more than three particles, to understand whether the evolution of a EH does map it into a single class or into a set of classes and to deal with protocols which use multipartite entanglement, which now has been proposed as a resource for QKD protocols for example in [6].

References

Quantum Machine Group Learning

Carlos Tavares
High-Assurance Software Laboratory/INESC TEC, Braga, Portugal,
ctavares@inesctec.pt

Abstract. Quantum machine learning has been gaining the attention of the communities of quantum computation and machine learning, as quantum computers seem able to provide significant advancements to machine learning. While the amount and diversity of techniques and algorithms available is already significant, they can be classified in two great groups: the ones that are purely quantum, having no classical analogue and the ones that are only enhancements of classical techniques, using quantum primitive algorithms. Concerning the latter group of techniques, it is observable that there is still a lot that can be imported from classical machine learning, for instance, the use of representation learning algorithms as data preparation steps for classification/regression techniques, which has been shown as effective in certain kinds of problems in fields such as computer vision. We aim to explore the possibility of application of these techniques in the quantum setting during this work.

Keywords: quantum machine learning, representation learning, hidden subgroup problem

1 Introduction

During the last years quantum machine learning became an hot topic of interest in the crossroads of machine learning and quantum computation, as a possible natural, and desirable, application for quantum algorithms. As surveyed in [3] the developments in the area came from different directions: from the creation of purely quantum techniques to the adaptation of the classical machine learning techniques to the quantum setting. Concerning the latter approach, many of the common techniques of classical computing have already been adapted to the quantum side, enhanced by quantum algorithms such as the Grover search algorithm [6], phase estimation, amplitude amplification, [4], and the HHL algorithm [8] for solving systems of linear equations. Successful examples range from clustering techniques in an unsupervised learning setting, principal component analysis for feature selection, as well as support vector machines [12], and neural networks [14], in a supervised learning setting.

Furthermore, there is a rich set of applications of group theory for machine learning in the classical setting, as it is observable, for instance, in the work of Kondor [9]. Groups are used to represent an algebra for the construction of symmetric transformations, e.g. transformations that leave an object unaltered.
One application for this lies in computer vision algorithms, where groups can be used to represent geometric transformations that do not deform objects (Abelian groups), such as rotation, translation, or scaling [2] transformations. The application of Fourier transform to data with an Abelian group structure has the effect of reducing it to its irreducible representations, hence finding the smallest amount of information that can represent the group.

Another more complex example is permutation learning, which consists in the identification of objects under the action of a permutation group, which has applications in ranking problems. Broadly speaking, group/representation theory has the effect of data/feature reduction, providing a possible preparation step to be used before the application of classification/regression techniques.

Quantum computers are exponentially more efficient in dealing with groups than classical computers as it can be observed, for instance, in the Shor algorithm. The algorithm works efficiently up to all normal subgroups [7], however, relevant groups to machine learning, such as the Dihedral [10] or symmetry groups [11], seem to too complex even for quantum computers. However, it is not known if quantum computers can provide better and faster approximations than classical computers, even for the most complex groups.

Hence, there seems to be a rich field of application for the use of the existent quantum algorithms for groups to be used as data preparation operations, before the application of quantum regression techniques, as similarly to what happens in the classical world. We aim to explore this possibility during this work.

References