Machine Learning in a Quantum World

Esma Aïmeur, Gilles Brassard, and Sébastien Gambs

Université de Montréal
Département d'informatique et de recherche opérationnelle
C.P. 6128, Succ. Centre-Ville, Montréal (Québec), H3C 3J7 Canada
{aimeur, brassard, gambsseb}@iro.umontreal.ca
http://www.iro.umontreal.ca/∼{aimeur, brassard, gambsseb}

Abstract. Quantum Information Processing (QIP) performs wonders in a world that obeys the laws of quantum mechanics, whereas Machine Learning (ML) is generally assumed to be done in a classical world. We initiate an investigation of the encounter of ML with QIP by defining and studying novel learning tasks that correspond to Machine Learning in a world in which the information is fundamentally quantum mechanical. We shall see that this paradigm shift has a profound impact on the learning process and that our classical intuition is often challenged.

1 Introduction

Quantum Information Processing (QIP) is the field that studies *the implication of quantum mechanics for information processing purposes*. Quantum information is very different from its classical counterpart: it cannot be measured reliably and it is disturbed by observation, but it can exist in a superposition of classical states. Classical and quantum information can be used together to realize wonders that are out of reach of classical information processing, such as being able to factorize efficiently large integers [21], search in an unstructured database with a quadratic speedup compared to the best classical algorithms [8] and allow two people to communicate in perfect secrecy under the nose of an eavesdropper having at her disposal unlimited computing power and technology [3].

Machine Learning (ML) is the field that studies techniques to *give to machines the ability to learn from past experience*. Typical tasks in ML include the ability to predict the class (*classification*) or some unobserved characteristic (*regression*) of an object based on some observations in *supervised learning*, or the ability to find some structure hidden within data (*clustering* or *density estimation*) in *unsupervised learning*. In general in ML, a machine is trained using a learning algorithm that takes as input a training dataset. This training dataset is implicitly assumed to be fundamentally classical, meaning that it contains "classical" observations about "classical" objects.

In this paper, we address the following question: What if the training dataset contains *quantum objects*? In particular what are the consequences for the learning process if we want to find analogues or develop new ML algorithms in this setting? The outline of the paper is as follows. In Section 2, we review some

L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 431–442, 2006.
© Springer-Verlag Berlin Heidelberg 2006

basic notions of QIP. In Section 3, we describe some previous encounters of ML and QIP before defining in Section 4 what learning could mean in the quantum context. We then illustrate our model in Section 5 by giving specific examples of clustering algorithms adapted to this new paradigm, including a simulated experiment. We conclude and state open problems in Section 6.

2 Review of Quantum Information Processing Concepts

In this section, we briefly review some essential notions of QIP [14]. A *qubit* (or *quantum bit*) is the quantum analogue of the classical bit. In contrast with its classical counterpart, a qubit can exist in a *superposition* of states. For instance, an electron can be *simultaneously* on two different orbits in some atom. Formally, using the Dirac notation, a qubit is described as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ where α and β are complex numbers called the *amplitudes* of classical states $|0\rangle$ and $|1\rangle$, respectively, subject to the *normalization* condition that $|\alpha|^2 + |\beta|^2 = 1$. When state $|\psi\rangle$ is *measured*, either $|0\rangle$ or $|1\rangle$ is observed, with probability $|\alpha|^2$ or $|\beta|^2$, respectively. Furthermore, measurements are *irreversible* because the state of the system *collapses* to whichever value ($|0\rangle$ or $|1\rangle$) has been observed, thus losing all memory of former amplitudes α and β.

All other operations allowed by quantum mechanics are *reversible* (and even *unitary*). They are represented by *gates*, much as in a classical circuit. For instance, the *Walsh–Hadamard* gate H maps $|0\rangle$ to $\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ and $|1\rangle$ to $\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle$. Figure 1 illustrates the notions seen so far, where time flows from left to right. Note that a *single line* carries quantum information, whereas a *double line* carries classical information; $\mathcal{M}$ denotes a measurement.

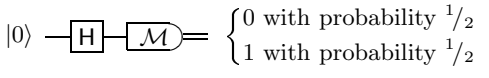

Fig. 1. Example of a simple quantum circuit

In this very simple example, we apply a Walsh–Hadamard gate to state $|0\rangle$, which yields $\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$. The measurement produces either 0 or 1, each with probability $|\frac{1}{\sqrt{2}}|^2 = 1/2$, and the state collapses to the observed classical value.

The notion of qubit has a natural extension, which is the *quantum register*. A quantum register $|\psi\rangle$, composed of n qubits, lives in a 2^n-dimensional Hilbert space $\mathcal{H}$. Register $|\psi\rangle = \sum_{i=0}^{2^n-1} \alpha_i|i\rangle$ is specified by complex numbers α_0, α_1,..., α_{2^n-1} subject to normalization condition $\sum |\alpha_i|^2 = 1$. Here, basis state $|i\rangle$ denotes the binary encoding of integer i. The *tensor product* $\otimes$ is used to represent the *composition* of two quantum systems. For instance, if we have two quantum states $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and $|\phi\rangle = \gamma|0\rangle + \delta|1\rangle$ and we put them next to each other, we can describe the composite system as $|\Gamma\rangle = |\psi\rangle \otimes |\phi\rangle = \alpha\gamma|00\rangle + \alpha\delta|01\rangle + \beta\gamma|10\rangle + \beta\delta|11\rangle$. As a shorthand notation, we write $|\psi\rangle^{\otimes k}$ for a quantum register composed of k identical copies of state $|\psi\rangle$.

Fig. 2. SWAP gate

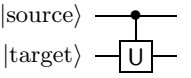

Fig. 3. Control-U gate

Unitary operations can also be applied to two or more qubits. For example, Fig. 2 illustrates a SWAP gate, which exchanges the qubits on its wires. Another kind of unitary gate frequently encountered in QIP is the *Control*-U gate, illustrated in Fig. 3: operation U is performed on the bottom wire, known as the *target*, if and only if the top wire, known as the *source*, is set to $|1\rangle$.

The *density matrix* is a formalism used to represent *one's knowledge about a particular quantum system*. It is a *complete* description of what can be observed about it. If we know that a quantum system is in a specific state $|\psi\rangle$, then its density matrix ρ is equal to $|\psi\rangle\langle\psi|$, where $\langle\psi|$ is the *conjugate transpose* of $|\psi\rangle$, the latter being considered as a column vector in the Hilbert space. In this case, ρ is said to be *pure*. A *mixed* state is a probability distribution over an ensemble $\{(p_1, |\psi_1\rangle), \ldots, (p_k, |\psi_k\rangle)\}$ of pure states, subject to $\sum_{i=1}^{k} p_i = 1$, where p_i is the probability associated with pure state $|\psi_i\rangle$. The density matrix ρ corresponding to this mixture is defined as $\sum_{i=1}^{k} p_i |\psi_i\rangle\langle\psi_i|$.

Two fundamental theorems set limits on what can be done with a quantum state. The *no-cloning theorem* [23] prevents us from cloning perfectly an unknown (or partially known) quantum state unless it is known to belong to a set of pairwise orthogonal states. A consequence of *Holevo's Theorem* [12] states that it is impossible to extract more than n bits of classical information from n qubits. Therefore, although n qubits need an exponential number 2^n of amplitudes to be described, only a linear amount of information can be extracted from them. Quantum information features many additional intriguing non-classical properties, such as *entanglement* and *interference*. See [14].

3 Previous Encounters of Machine Learning with Quantum Information Processing

In any field of computer science, it is natural to ask whether or not it is possible, using the QIP paradigm, to obtain more efficient and more powerful information processing capabilities. For example, we can seek faster algorithms, or savings on the communication cost in distributed contexts, or security upgrading in cryptographic settings, etc. When looking specifically at ML and QIP, there are several ways one could imagine to try mixing them. From a theoretical point of view, some work has already been done in computational learning theory that compares learnability in the classical and the quantum settings. Two models have been generalized to the quantum world: the Probably Approximately Correct (PAC) learning model of Valiant [22] and the model of exact learning from membership queries of Angluin [1].

According to the quantum analogues of these models, the goal is to infer properties of a function f, whose access is provided through a quantum *oracle*

that can be queried in a superposition of questions. Servedio and Gortler have studied both Valiant's and Angluin's models and proved interesting results [20]. One of their discoveries is that the quantum paradigm does *not* provide any advantage in exact learning from membership queries or in PAC learning, if we care only about the number of queries: every function (concept) that can be "learned" (in the specific sense given in each of these models) with a polynomial number of quantum queries can also be learned with a polynomial number of classical queries. However, this equivalence does not hold for *computing time*: Servedio [19] has constructed an explicit class of concepts that are *polynomial-time learnable from quantum membership queries but not from classical queries*, based on the cryptographic assumption that one-way functions exist.

Other previous encounters of ML with QIP include the definition of quantum analogues for ML approaches such as neural networks [6], the design of classical clustering algorithms inspired from quantum mechanics [13], the application of the maximum likelihood principle to quantum channel modelling [24], etc.

4 Learning in a Quantum World

4.1 Training with a Quantum Dataset

Machine learning algorithms learn from a training dataset, which is given as input to the algorithm. The dataset contains observations about objects, which were obtained empirically or acquired from experts. In the *classical* setting, the observations and the objects are implicitly considered to be classical. In the case of *supervised* learning, the training dataset can be described as $D_n = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i$ would be some *observations* on the *characteristics* of the i^{th} object (or data point) and y_i is the corresponding *class* of that object. As a typical example, each object can be described using d real-valued attributes and we are dealing with binary classification. In this case $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$. The main difference between supervised and *unsupervised* learning is that in the latter case the y_i values are unknown. This could mean that we know the possible labels in general but not the *specific* label of each data point, or that even the number of classes and their labels are unknown to us.

In a quantum world, a ML algorithm still needs a training dataset from which to perform learning, but this dataset now contains *quantum* objects. This forces us to rethink the entire learning process because quantum information obeys different physical laws, compared to classical information. Quantum mechanically, a training dataset containing n quantum states can be described as $D_n = \{(|\psi_1\rangle, y_1), \ldots, (|\psi_n\rangle, y_n)\}$, where $|\psi_i\rangle$ is the i^{th} quantum state of the training set and y_i is the corresponding class of that quantum state. A typical example occurs when a quantum state lives in a Hilbert space formed by d qubits: $|\psi_i\rangle \in \mathbb{C}^{2^d}$ and $y_i \in \{-1, +1\}$ for binary classification. Even though we restrict ourselves to classical classes in this work, further generalization would be possible in which objects can be in a quantum superposition of classes.

Imagine a scenario in which a space probe has been sent away from Earth. In its exploration process, the probe encounters various quantum phenomena,

from which it samples to constitute a training dataset. The samples could be labelled by the origin of the quantum phenomenon (*supervised learning*), or left unlabelled if their source is unknown (*unsupervised learning*). Afterwards, the probe would search for an intrinsic hidden structure within the quantum data.

4.2 Learning Classes

There are many ways in which quantum states can be specified in the training dataset. To this end, we introduce several learning classes that differ in the form of the training dataset, the learner's technological sophistication and his learning goal. For learning class $L_{goal}^{context}$, subscript *goal* refers to the learning goal and superscript *context* to the training dataset and/or the learner's abilities. Possible values for *goal* are c, which stands for doing ML with a classical purpose, and q for ML with a quantum purpose. The superscript *context* can be c for "classical" if everything is classical (with a possible exception for the goal) or q if something "quantum" is going on. Other values for *context* can be used when we need to be more specific. For example, L_c^c corresponds to ML in the usual sense, in which we want to use classical means to learn from classical observations about classical objects. Another example is L_c^q, in which we have access to a quantum computer for help but the goal remains to perform a classical ML task: the quantum computer could serve to speed up the ML process.

We are more concerned with the case of "$goal = q$". The simplest instance corresponds to L_q^c, which is defined as the quantum learning class in which all the training set quantum state descriptions are given classically (i.e. $D_n = \{(\psi_1, y_1), \ldots, (\psi_n, y_n)\}$, where ψ_i is the classical description of quantum state $|\psi_i\rangle$). Learning becomes more challenging when the dataset is available only in its quantum form, in which case more copies make life easier. Class $L_q^{\otimes k}$ is defined as the learning class in which we are given k copies of each training quantum state (i.e. $D_n = \{(|\psi_1\rangle^{\otimes k}, y_1), \ldots, (|\psi_n\rangle^{\otimes k}, y_n)\}$; recall that $|\psi_i\rangle^{\otimes k}$ symbolizes k copies of state $|\psi_i\rangle$). Contrast these classes with ML in a classical world (such as L_c^c), in which additional copies of an object are obviously useless.

4.3 Possible Learning Strategies

Several types of strategies can be imagined, depending on the learning class and the task we wish to realize. Here, we study the case of *binary* classification for the purpose of illustrating the strategies with a "concrete" example. Consider the quantum classification task of predicting the class of an unknown quantum state $|\psi_?\rangle$ *given a single copy of this state* (This constraint can be relaxed by considering the case of multiple copies of the state to be classified—see the *quantum template matching* problem of [17]). The easiest situation occurs when $D_n \in L_q^c$ since we have complete classical knowledge of the *training* states. Despite this advantage, it is not possible in general to devise a process that would always classify correctly unknown quantum states even should they be known to be identical to one of the training states (see last paragraph in Section 4.5). Nevertheless, when $D_n \in L_q^c$, it is possible to analytically devise the optimal measurement that minimizes the training error. It remains to be seen how well such an approach

would fare when faced with a new quantum state. Alternatively, in some cases, we can devise an *unambiguous discrimination* measurement, which would never give the wrong classification for objects in the training set, but would sometimes refuse to answer at all. This is analogous to the classical case in which a classifier has to either predict class "−1" or "+1" with the highest possible confidence, but is allowed to abstain (by answering "0") when it has low confidence in its best prediction.

If $D_n \in L_q^{\otimes k}$ then possible strategies include: (1) estimate the training set quantum states by making measurements (joint or not) on some copies; (2) devise a classification mechanism that uses copies of the training set quantum states only at the time of demand (i.e. when the time to classify $|\psi_?\rangle$ comes); or (3) compose a hybrid strategy based on (1) and (2).

Note that strategy (1) corresponds to the concept of *quantum tomography*, in which one tries to reconstruct as faithfully as possible the classical description of a quantum state from a finite number of copies of this state. Strategy (2) is unique to the quantum world and we call a classifier resulting from this strategy a *one-time classifier* because we sacrifice some parts of the training set when the classification need arises. Using this strategy, once the classification is done, the information contained in the sacrificed part of the training set is lost forever. This has no classical counterpart because nothing in principle prevents a classical classifier from being used an unbounded number of times. A hybrid strategy (3) could be built on the advantages of both previous strategies, for instance by acquiring some classical knowledge about the training set by performing state estimation and then using this knowledge in the design of a one-time classifier.

We have considered in this paper only the case of binary classification. Moving to the multiclass setting has to be done with care [18], even more so if we consider quantum classes that can be in a superposition.

4.4 Hierarchy of Quantum Learning Classes

The different quantum learning classes form a hierarchy in an information-theoretic sense, meaning that the higher a class is located in the hierarchy, the potentially better classification of an unknown state it allows. The class L_q^c is at the top of the hierarchy since it corresponds to *complete knowledge* about the quantum states in the training set. Let $\equiv_\ell$, $\leq_\ell$ and $<_\ell$ denote the *equivalence*, the *weaker or equal* and the *strictly weaker* relationships within the hierarchy, respectively. The following statements are obvious.

- $L_q^{\otimes k} \equiv_\ell L_q^c$ as $k \to \infty$.
- $L_q^{\otimes 1} \leq_\ell \ldots \leq_\ell L_q^{\otimes k} \leq_\ell L_q^{\otimes k+1} \leq_\ell \ldots \leq_\ell L_q^c$.
- $L_q^{\otimes k} + L_q^{\otimes 1} \leq_\ell L_q^{\otimes k+1}$, where "+" denotes a restriction that the first k copies must be measured separately from the the last.

The interesting question is whether or not this hierarchy is strict: can all these $\leq_\ell$ be replaced by $<_\ell$? There are good reasons to believe that the answer is positive since it is usually the case that more information can be obtained about

a quantum state when more copies are available, and it has been proven in some cases that joint measurements are more informative than individual measurements [15, 5]. But it does not *necessarily* follow that this additional classical information provides for a better quantum classifier when the time comes to identify unknown state $|\psi_?\rangle$.

4.5 Bounds on the Training Error

Let m_- be the number of quantum states in D_n for which $y_i = -1$ (negative class) and its complement m_+ be the number of states in D_n for which $y_i = +1$ (positive class), with $m_- + m_+ = n$, the total number of data points in D_n. The mixture ρ_- is defined as $\frac{1}{m_-}\sum_{i=1}^{n}\frac{1-y_i}{2}|\psi_i\rangle\langle\psi_i|$ and the mixture ρ_+ as $\frac{1}{m_+}\sum_{i=1}^{n}\frac{1+y_i}{2}|\psi_i\rangle\langle\psi_i|$. If $D_n \in L_q^c$, the problem of classifying an unknown state $|\psi_?\rangle$ taken from the training set is equivalent to distinguishing between ρ_- and ρ_+. The success probability of this classification process is linked to the *statistical overlap* of these distributions. In fact, this kind of problem has already been studied in *quantum detection and estimation theory* [10], a field that predates QIP itself. Some results from this field can be used to *give bounds on the best training error* we could hope for from a ML algorithm. For instance, the probability of distinguishing between the two classes with the optimal quantum process is bounded above by $(1 + D(\rho_-, \rho_+))/2$, where $D(\rho_-, \rho_+) = \mathrm{Tr}|p_-\rho_- - p_+\rho_+|$ is a distance measure between ρ_- and ρ_+ due to Helstrom [10] (here, p_- and p_+ represent the *a priori* probabilities of ρ_- and ρ_+, respectively). For unambiguous discrimination, bounds have been developed much more recently [11]. The goal of a ML algorithm in the quantum setting is to give a constructive way to come close to (or to achieve) these bounds.

Note that, contrary to classical ML, where it is always possible—albeit not always advisable—to bring the training error down to zero (for instance using a memory-based classifier such as 1-nearest neighbour), it is *impossible to do so in the quantum case* unless the states of the training set are pairwise orthogonal.

5 Illustration: Clustering with a Quantum Dataset

5.1 Measure of Distance Between Quantum States

The quantity $Fid(|\psi\rangle, |\phi\rangle) = |\langle\phi|\psi\rangle|^2$ is an important notion in QIP, which is called the *fidelity*. Note that the fidelity is similar to a measure commonly used in *classical* information retrieval, namely the *cosine similarity*. The fidelity is a *similarity measure* between two quantum states, which ranges from 0 if they are orthogonal (meaning perfectly distinguishable) to 1 if the states are identical. Properties of the fidelity [14, §9.2.2] include *symmetry*, $Fid(|\psi\rangle, |\phi\rangle) = Fid(|\phi\rangle, |\psi\rangle)$, and *invariance under unitary operations*, meaning that applying the same unitary operation U on two different quantum states does not change their fidelity: $Fid(|\mathsf{U}\psi\rangle, |\mathsf{U}\phi\rangle) = Fid(|\psi\rangle, |\phi\rangle)$. In its standard form, the fidelity is not really a metric because it does not obey the triangle inequality, but it can be made to do so if we use $Dist(|\psi\rangle, |\phi\rangle) = \arccos Fid(|\psi\rangle, |\phi\rangle)$.

The value of $Dist(|\psi\rangle, |\phi\rangle)$ ranges from 0 if the states are identical to $\frac{\pi}{2}$ if they are orthogonal. This *distance* now respects the triangle inequality $Dist(|\psi\rangle, |\phi\rangle) \leq Dist(|\psi\rangle, |\varphi\rangle) + Dist(|\varphi\rangle, |\phi\rangle)$.

5.2 Control-Swap as a Fidelity Estimator

The Control-Swap test (C-Swap test) [2, 4] makes it possible to estimate the similarity between two unknown quantum states $|\psi\rangle$ and $|\phi\rangle$, as measured by their fidelity $Fid(|\psi\rangle, |\phi\rangle)$. Figure 4 illustrates this concept.

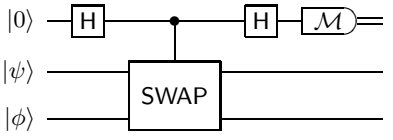

Fig. 4. Circuit of the Control-SWAP test

The input to the circuit is $|0\rangle|\psi\rangle|\phi\rangle$. After applying the first Walsh–Hadamard gate H, the state evolves to the superposition $\frac{1}{\sqrt{2}}|0\rangle|\psi\rangle|\phi\rangle + \frac{1}{\sqrt{2}}|1\rangle|\psi\rangle|\phi\rangle$. Application of the C-Swap gate, exchanges $|\psi\rangle$ and $|\phi\rangle$ if the state of the upper wire is $|1\rangle$. Therefore, the state evolves to $\frac{1}{\sqrt{2}}|0\rangle|\psi\rangle|\phi\rangle + \frac{1}{\sqrt{2}}|1\rangle|\phi\rangle|\psi\rangle$. Afterwards, the second Walsh–Hadamard gate H transforms the state to $\frac{1}{2}|0\rangle(|\psi\rangle|\phi\rangle + |\phi\rangle|\psi\rangle) + \frac{1}{2}|1\rangle(|\psi\rangle|\phi\rangle - |\phi\rangle|\psi\rangle)$. Finally, measurement of the top qubit yields classical outcome 0 with probability 1 if $|\psi\rangle$ and $|\phi\rangle$ are identical. In general, the measurement outcome is 1 with probability $\frac{1}{2} - \frac{1}{2}|\langle\phi|\psi\rangle|^2$. It follows that the C-Swap test provides an *estimator* for the *fidelity* between $|\psi\rangle$ and $|\phi\rangle$. With k copies of these states, we can estimate $Fid(|\psi\rangle, |\phi\rangle)$ as $1 - 2 \times \#|1\rangle/k$. Note that a side effect of the C-Swap test is to irreversibly disturb the input states, unless they happened to be identical.

5.3 Examples of Possible Quantum Clustering Algorithms

If $D_n \in L_q^{\otimes k}$, the most obvious strategy for clustering the training set of quantum states would follow the "type (1)" approach outlined in Section 4.3: By way of quantum tomography, use all the available copies in order to reconstruct a classical description of each training state. A classical clustering algorithm could then be applied on the resulting approximate descriptions. Obviously, the accuracy of this approach is limited by that of the classically estimated quantum states. Fortunately, this is a well-studied problem.

The optimal mean fidelity achievable for the reconstructed state $|\psi_{guess}\rangle$, compared to the true state $|\psi_{true}\rangle$, is a function of the dimension of the state and of the number of copies. The exact formula is $Fid(|\psi_{guess}\rangle, |\psi_{true}\rangle) = \frac{k+1}{k+d}$, where k is the number of available copies and d the dimension of the Hilbert space (see [9] for an interesting discussion on the subject). This means that in order to achieve good fidelity in the reconstruction, an exponential number of copies in the number of qubits is required. In the case when d is small and k is large, this

tomographic approach can generate a reasonably faithful classical description of the state. From this classical description, it is possible to compute directly (i.e. using a classical computer), the fidelity between two quantum states. From there, classical clustering algorithms can be used.

Recall that the main goal of clustering is to *group similar objects together* while *putting objects that are dissimilar into different clusters*. Thus, it should be intuitively clear that the approach to clustering outlined above is wasteful of the precious quantum resources. Indeed, there is no need to determine a classical approximation of two quantum states if the genuine purpose of the operation is merely to estimate the distance between them according to some metric. A more promising approach to quantum clustering is to estimate that distance directly through a joint measurement of the two quantum states. The simplest way to do this (but not necessarily the best) is to use the C-Swap test of Section 5.2 to estimate fidelity as a measure of distance.

For example, if we are in $L_q^{\otimes(n-1)e}$, we can estimate the fidelity between each pair of states in the training dataset by applying the C-Swap test e times independently for each pair. We could then use a classical algorithm, such as *k-medians*, to perform clustering. In the next section, we report on a simulated experiment along these lines.

Other strategies can be devised, which are even more quantum. For instance, we could adapt a classical algorithm to the quantum setting, such as an *agglomerative algorithm* that would grow clusters around *quantum seeds* in an adaptive manner. This algorithm first sacrifices some parts of the training set during the seeding phase to identify states whose pairwise dissimilarity is high, which will be used as seeds. During the second phase, each state is compared to the seeds using the C-Swap test estimation and agglomerated around the most similar one.

5.4 Experimentation

In this section, we present some preliminary results on a very simple clustering experiment. As a proof of concept, we chose to test on synthetic data an algorithm based on the C-Swap test outlined in the previous section. The data is composed of five clusters. Each cluster is centred on a 13-qubit pure state generated randomly and uniformly according to the Haar measure in a 8192-dimensional Hilbert space. For each cluster, twenty pure states were obtained by applying a random perturbation to the cluster centre so that the fidelity of the resulting states with the cluster centre were never below some threshold, henceforth called the *fidelity threshold*. Note that we can make the clustering problem more difficult by lowering this threshold since this results in less dense clusters, which can even overlap when the fidelity threshold is too low.

We used the C-Swap test to estimate the fidelity between each pair of states in the training set. The greater the number of samples are available for the C-Swap test, the more accurate is the resulting estimate. However, there is no need to estimate those fidelities with exceedingly high precision (which would be too expensive in the required number of copies) because a rough estimate of the

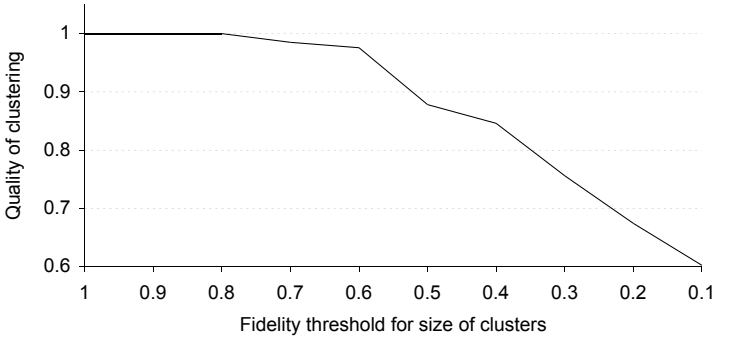

Fig. 5. Evolution of the clustering quality (averaged over 5 trials) as the fidelity threshold decreases. A quality value of 1 indicates perfect clustering. Random "clustering" would correspond to a quality of approximately 0.37, below this curve.

similarity between states is generally sufficient in order to group similar objects in the same cluster and to put dissimilar objects in different clusters.

After generating the 100 pure states as described above, we permuted them randomly and we removed all information as to which was their initial cluster. The first clustering phase consisted in building a *similarity matrix* in which each entry (x, y) represented the estimated fidelity between states x and y. The fidelity between each pair of states was estimated using the C-Swap test with 20 samples, which corresponds roughly to an absolute error of 0.15 in the fidelity estimates. During the second clustering phase, the similarity matrix was given as input to a classical clustering algorithm inspired by k-medians. This algorithm selected five states at random as cluster centres (we told the algorithm the original number of clusters) and then performed iteratively two steps until stabilization (or quasi-stabilization). In the first step, states were assigned to their most similar centre; in the second step, the centre of each cluster was recomputed by selecting the point that maximized its similarity with the other points of the cluster.

The *quality* of this clustering process was measured by determining how well points that were originally in the same cluster were still together and conversely how well points that were originally in different clusters remained separated. For this purpose, we used a metric ranging from 1 if all the points are perfectly placed after clustering to -1 if they are completely messed up. According to this metric, a completely random "clustering", in which each state would be assigned to a random cluster, exhibits an average "quality" of approximately 37%.

Figure 5 summarizes the results obtained for different values of the fidelity threshold. As expected, the quality of the clustering process is at its best when the fidelity threshold is high, because in this case the original clusters form tightly packed balls that are far from each other. This quality remains above 97% even for a fidelity threshold as low as 0.6. The quality continues to decrease with the fidelity threshold, reaching a level of roughly 60% for a fidelity threshold of 0.1. Note that this is not very good compared to the worthless quality of 37% obtained by random "clustering". This phenomenon is due partly to the fact

that clusters start to overlap as the fidelity threshold decreases, and also to the imperfect quality of the estimates obtained using the C-Swap test. Naturally, if one is not satisfied with those C-Swap estimates, there is always the option of sacrificing more copies of the training states to refine the estimates.

6 Conclusions and Open Problems

In this paper, we have described a novel task and a new learning model that corresponds to performing ML in a world in which information behaves according to the laws of quantum mechanics—which is, of course, *the* real world. Using quantum information has a great impact on the learning process and many of our classical ML intuitions are being challenged in this world. On the other hand, this model offers many interesting questions and perspectives, and studying it could lead to insights both in ML and QIP. For practical purposes, quantum ML algorithms have the potential to give a constructive answer to some detection scenarios. From a theoretical viewpoint, studying this model could lead to discoveries about the notion of learning from a generic perspective and how it is linked to the underlying physical theory on which it is based.

We have illustrated our model by showing some examples of clustering algorithms, but algorithms for other ML tasks can also been designed. Currently, we are developing quantum versions of ID3 [16] and AdaBoost [7] for classification by exploiting, respectively, the relationship between Shannon and von Neumann entropy, and the resemblance between the notions of weak classifier and weak measurement. We conclude this paper by mentioning but a few of our many open problems: define analogues of classical notions of ML to the quantum setting such as the *test* and the *generalization errors* or the *margin*, study the different models of classical and quantum *noise* (see [14, §8.3] for different forms of quantum noise) and how they affect the robustness of the quantum ML algorithms, improve classical ML algorithms with quantum computation or devise brand new algorithms adapted to this setting, study and define other classes of learning such as when the training set is composed of mixed states, etc.

Acknowledgements

The authors are supported in parts by the Natural Sciences and Engineering Research Council of Canada. Gilles Brassard is also supported by the Canadian Institute for Advanced Research and the Canada Research Chair Programme.

References

1. Angluin, D., "Queries and concept learning", *Machine Learning* **2**, pp. 319–342, 1988.
2. Barenco, A., Berthiaume, A., Deutsch, D., Ekert, A., Jozsa, R. and Macchiavello, C., "Stabilisation of quantum computations by symmetrisation", *SIAM Journal of Computing* **26**(5), pp. 1541–1557, 1997.

3. Bennett, C. H. and Brassard, G., "Quantum cryptography: Public key distribution and coin tossing", *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, Bangalore, pp. 175–179, 1984.
4. Buhrman, H., Cleve, R., Watrous, J. and de Wolf, R., "Quantum fingerprinting", *Physical Reviews Letters* **87**(16), article 167902, 2001.
5. Chiribella, G., Mauro D'Ariano, G., Perinotti, P. and Sacchi, M., "Covariant quantum measurements which maximize the likelihood", *Physical Reviews A* **70**, article 061205, 2004.
6. Ezhov, A. A. and Berman, G. P., *Introduction to Quantum Neural Technologies*, Rinton Press, 2003.
7. Freund, Y. and Shapire, R., "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences* **55**(1), pp. 119–139, 1997.
8. Grover, L. K., "Quantum mechanics helps in searching for a needle in a haystack", *Physical Review Letters* **79**, pp. 325–328, 1997.
9. Hayashi, A., Hashimoto, T. and Horibe, M., "Reexamination of optimal quantum state estimation of pure states", *Physical Reviews A* **72**, article 032325, 2005.
10. Helstrom, C. W., *Quantum Detection and Estimation Theory*, Academic Press, 1976.
11. Herzog, U. and Bergou, J. A., "Optimal unambiguous discrimination of two mixed quantum states", *Physical Reviews A* **71**, article 050301, 2005.
12. Holevo, A. S., "Bounds for the quantity of information transmitted by a quantum mechanical channel", *Problems of Information Transmissions* **9**, pp. 177–183, 1973.
13. Horn, D. and Gottlieb, A., "The method of quantum clustering", *Proceedings of the Conference Neural Information Processing Systems (NIPS)*, pp. 769–776, 2001.
14. Nielsen, M. A. and Chuang, I. L., *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
15. Peres, A. and Wootters, W. K., "Optimal detection of quantum information", *Physical Reviews Letters* **66**(9), pp. 1119–1122, 1991.
16. Quinlan, J. R., "Induction of decision trees", *Machine Learning* **1**, pp. 81–106, 1986.
17. Sasaki, M. and Carlini, A., "Quantum learning and universal quantum matching machine", *Physical Reviews A* **66**(2), article 022303, 2002.
18. Sen, P., "Random measurement bases, quantum state distinction and applications to the hidden subgroup problem", *Proceedings of 21st Annual IEEE Conference on Computational Complexity (CCC)*, to appear, 2006.
19. Servedio, R., "Separating quantum and classical learning", *Proceedings of the International Conference on Automata, Languages and Programming (ICALP)*, pp. 1065–1080, 2001.
20. Servedio, R. and Gortler, A., "Equivalences and separations between quantum and classical learnability", *SIAM Journal of Computing* **33**(5), pp. 1067–1092, 2004.
21. Shor, P. W., "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", *SIAM Journal of Computing* **26**, pp. 1484–1509, 1997.
22. Valiant, L. G., "A theory of the learnable", *Communications of the ACM* **27**, pp. 1134–1142, 1984.
23. Wootters, W. K. and Żurek, W. C., "A single quantum cannot be cloned", *Nature* **66**, pp. 802–803, 1982.
24. Ziman, M., Plesch, M., Bužek, V. and Štelmachovič, P., "Process reconstruction: From unphysical to physical maps via maximum likelihood", *Physical Reviews A* **71**, article 022106, 2005.