Classical and Quantum Information

Michael Spence[1, *]

[1] *Quantum Information Final Project*

(Dated: January 18, 2018)

I briefly outline the basic principles of classical and quantum information theory before contrasting the two. I include a brief describtion of Grover's data base search algorithm, comparing with classical search algorithms.

PACS numbers:

INTRODUCTION

Generally speaking, information theory can be considered as the study of information quantization, communication, storage and compression. The modern view of information had it's origin in the 1948 paper by Claude E. Shannon, "A Mathematical Theory of Communication." One of the largest conceptual advances of the paper was to consider information as the uncertainty associated with the result of a random variable. This can be quantized in the form of Shannon's entropy, thereby opening many previously inaccesible problems to mathematical analysis.[1] The paper was hugley influential, and with the rapid advance of computers these methods found plenty of scope for use. Modern cryptography, signal processing and even the internet all owe their existence in part to concepts and methods outlined in Shannon's paper and later devolped by others.

The fundamentals of quantum mechanics were developed over the first three decades of the twentieth century, but suprislingly it took until the 1980's for researchers to start seriously considering the possibility of exploitng quantum phenomena to manipulate information.[2] One of the first steps was due to Paul Benioff, who demostrated that a system of spin $\frac{1}{2}$ particles can emulate the operations of a Turing machine. This shows that a quantum mechanical system can be at least as efficent as a classical computer, if not better.[3] Over the next two decades examples of quantum algorithms were found which vastly outperform their classical counterparts. For example Shor's algorithm for number factorization, and Grover's algorithm for database search. These algorithms take advantage of entanglement, a phenomenon not present in classical physics and impossible to emulate on a classical computer. Many of the potential applications of quantum information theory remain unreachable due to the immense challenges involved in constructing a quantum computer. It has however found uses in some fields, notably quantum cryptography, which, unlike classical cryptography, has the potential to produce truly unbreakable codes.[5]

To make similarities and differnces between quantum and classical information more apparent it is neccesary to go into greater detail.

CLASSICAL INFORMATION

For a random variable X with outcomes x_i, and corresponding probabilities p_i we define the Shannon entropy as,

$$H(X) = -\sum_i p_i log(p_i). \tag{1}$$

This is the fundamental measure of information in classical information theory. This definition turns out to be the only option for a function measuring information, assuming certain requirements of intuition, such as positive definitness.[1] There is a degree of freedom in the base of the log, the most common choice being two. The baisc unit of classical information corresponds to a random variable with two equiprobable outcomes. This is called a "bit", and working in log_2 we can deduce it has entropy 1.

The Shannon entropy satisfies several identities which have intuitive interpretations. For example, for the information stored in two random variables, X and Y, we have,

$$H(X,Y) = H(X) + H(Y|X), \tag{2}$$

where $H(Y|X) = -\sum_{x,y} p(x,y)log(p(y|x))$. This can be thought of as saying that the information gained from learning the outcome of X and Y is equal to the information gained by learning the outcome of X, and then from learning the outcome of Y, given that you already know X. Another property of Shannon entropy is that conditioning never increases information.

$$H(X|Y) \leq H(X), \forall X, Y. \tag{3}$$

Again this is to be expected. Intuitivley learning about a system can only ever decrease our uncertainty.

An important concept in information theory is the mutual information between two random variable , X and Y, defined as

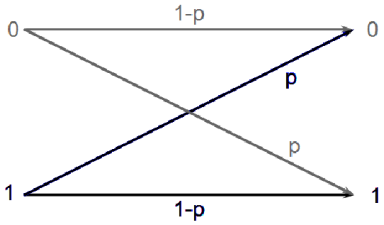
$$I(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{4}$$

This can be thought of as the information shared between the two variables, ie. the information lost from X upon

learning the outcome of Y, and vice versa. Mutual information turns out to be a particularly ubiquitous object in information theory. For example, consider a system designed to communicate the result of a random variable, X. The outcome of the procedure will be a random variable Y. Such a system is known as a channel. How can we measure the sucess of this channel? The answer is the mutual information of the variables of X and Y maximised over the possible probability distributions for X. This is called the channel capacity,

$$C = max_{p(x)} I(X : Y). \tag{5}$$

The simplest non-trivial example of a channel is one with two inputs and two outputs. Both inputs have an equal probability p of returning the correct result, and probability $1 - p$ of returning the incorrect result. This is called the symmetric binary channel. We represent this diagramiatically as



By setting the probability of X being 0 as q and then maximising over it, we find that $C = 1 - H(p)$, where $H(p) = -(p)log(p) - (1 - p)log(1 - p)$.

These are just a handful of aspects of classical information theory. Let's examine their quantum counterparts.

QUANTUM INFORMATION

For the purposes of quantum information we represent a system by a self adjoint, positive semi-definite operator of trace one, acting on the Hilbert space associated with the system, $\mathcal{H}$. This is known as a density operator and is usually written as ρ. There are several cases when the density operator has an obvious interpretation. For example if the density operator can be written in the form $\rho = |\psi\rangle \langle\psi|$, for some $|\psi\rangle \in \mathcal{H}$. This is interpreted as the observer knowing with ceratainty that the system is in the state $|\psi\rangle$. This is known as a pure state. A system in which the exact state is unknown is said to be in a mixed state. If for an orthonormal basis of $\mathcal{H}$, $|\phi_i\rangle$, we can write the density operator as

$$\rho = \sum_i p_i |\phi_i\rangle \langle\phi_i|, \tag{6}$$

then this can be interpreted as the system being in state $|\phi_i\rangle$ with probability p_i. In this case the density operator is an analouge of a classical random variable. This is due to the fact that mutually orhtogonal states are distinguishable by measurement, meaning that the system can only be in a certain number of non-overlapping states. By non-overlapping I mean that if you knew the system was in state $|\phi_i\rangle$, it would be impossible to then measure it being in state $|\phi_j\rangle$, for $i \neq j$. The same is not true for the states $|0\rangle$, and $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Despite the fact these states are distinct it is impossible to distinguish them with certainty through measurement.

We describe a composite of two systems with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, by a density operator acting on $\mathcal{H}_A \bigotimes \mathcal{H}_B$, usually denoted ρ_{AB}. To describe the state of one of the system from the perspective of an observer with acess to only a subsystem we trace over the other subsystem. For example $\rho_A = tr_B(\rho_{AB})$.

The fundamental measure of quantum information is known as von Neumann entropy. For a density operator ρ the von Neumann entropy is defined as,

$$S(\rho) = -tr(\rho log(\rho)). \tag{7}$$

Because the trace of a matrix is independent of basis the von Neumann entropy doesn't depend on how we represent the density operator. So for an arbitray denisty operator we have,

$$S(\rho) = -\sum_i \lambda_i log(\lambda_i), \tag{8}$$

where λ_i are the eigenvalues of the operator. It also vanishes for pure states, which is to be expected. There shouldn't be any uncertainty if you know exactly what state the system is in. The entropy of a composite syetm is defined as,

$$S(A, B) = -tr(\rho_{AB} log(\rho_{AB})). \tag{9}$$

In analouge to classical information The von Neumann entropy has many of the same properties as Shannon's entropy. Similarly to eq.2 we have,

$$S(A, B) = S(A) + S(B|A). \tag{10}$$

We also define mutual information in a similar way to eq.4.

$$S(A : B) = S(A) - S(A|B) = S(B) - S(B|A). \tag{11}$$

The fundamental component of quantum information is the "qubit", defined as a system with a two dimensional Hilbert space, a spin $\frac{1}{2}$ particle being the most common example. A pure qubit state needs two real numbers to be fully described. This can be parameterized by a 2-sphere known as the Bloch sphere.

Just as in classical information theory we have the concept of quantum channels. Generally this is a map between Hilbert spaces for two systems. The input is a density matrix acting on one Hilbert space, and the output is the density matrix created which acts on the other Hilbert space. One simple example is time evolution of a

system. In quantum mechanics states are evolved by the action of a unitary operator, ie. $|\psi\rangle \rightarrow U|\psi\rangle$. So for a denisty operator we have,

$$\rho \rightarrow U\rho U^{\dagger}. \tag{12}$$

This could be considered a means of transferring quantum information if two observers had acess to a system at different times. If the Hamiltonian of the system along with the time elapsed between observations was known, then the transformation could be inverted making it a noiseless channel.

QUANTUM VS CLASSICAL

Despite having many similarities, classical and quantum information have many important differences. Firstly consider the kind of uncertainty that is being measured in each case. For a classical system with zero Shannon entropy there is no uncertainty. Any observable aspect of the system is known or can be predicted with certainty. But consider a quantum system with zero von Neumann entropy, or in other words, in a pure state. Even though we know what quantum state the system is in there is still uncertainty present. Take a spin $\frac{1}{2}$ particle for example. If we know it is in the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$ with respect to the z-axis, then it has zero von Neumann entropy. But there is still uncertainty in the system. The result of a meausrement of the spin in the z direction could not be predicted with certainty. This is simply due to essential indeterminacy of quantum mechanics. Hence von Neumann entropy does not measure uncertainty in the intuitive sense the way Shannon's entropy does. A tentaive analogy can be drawn when we know that the state of the system is one of an ensemble of orthogonal states, $\{|\psi_i\rangle\}$, each occuring with probability $\{p_i\}$. In this case the density matrix can written in the same form as eq.6. Calculating the von Neumann entropy we find that it is equal to the Shannon entropy, if the potential state of the system is considered as a classical random variable.

$$S(\rho) = -\sum_i p_i log(p_i) = H(p_i). \tag{13}$$

Since Hermitian matricies can always be diagonalized any density operator can be written in this form. However we might not find it useful to think of the system in terms of the basis $|\psi_i\rangle$, so the correspondence is not always enlightening.

If we consider the case in which the system is potentially in a number of states, $\{|\phi_i\rangle\}$, which are not neccesarily orthogonal, with probabilities $\{p_i\}$, then we find that,

$$S(\rho) \leq H(p_i), \tag{14}$$

with equality being reached only when $\{|\phi_i\rangle\}$ are orthogonal. This can be interpreted as an ensemble of indistinguishable states having less uncertainty than orthogonal ones.

Another important difference between classical and quantum information is entanglement. Classically if we know about the overall state of a composite system then the uncertainty associated with the subsytems must be zero as well. This can be deduced from the fact Shannon's entropy is positive definite, even for conditional distributions.

$$\begin{gathered} H(A,B) = 0 = H(A) + H(B|A) = H(B) + H(A|B) \\ \implies H(A) = H(B) = 0. \end{gathered}$$

The same does not hold for quantum systems. Consider a system of two spin $\frac{1}{2}$ particles with density matrix

$$\rho_{AB} = \frac{1}{2}(|\uparrow\uparrow\rangle\langle\uparrow\uparrow| + |\uparrow\uparrow\rangle\langle\downarrow\downarrow| + |\downarrow\downarrow\rangle\langle\uparrow\uparrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|). \tag{15}$$

This is a pure state so we have that $S(\rho_{AB}) = 0$. But calculating the density matrices describing the subsystems we find

$$\rho_A = \rho_B = \frac{1}{2}(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|). \tag{16}$$

Using the eigenvaule formula , eq.8, and working in log_2, we find that $S(A) = S(B) = 1$. So despite the fact that the overall system has no entropy, the subsystems do. This is often interpreted as being due to the entanglement of the subsystems. Quantum infomation is not just dependent on the state of a system, but also how it related to others.

From the above example in conjunction with eq.10 we can also determine another divergence with classical information. For the above system we have $S(B|A) = -1$. Classically the conditional Shannon entropy is positive definite, as is to be expected. Any kind of information function, even conditional, would be expected to have zero as it's lower bound, representing complete knowledge of the system. We could view this as meaning that the uncertainty of B is increased upon finding out the state of A, but this is not terribly intuitive.

Both of the above are due in part to what is known as subadditivity, a property of von Neumann entropy. This states for two systems , A and B, we have,

$$\begin{gathered} S(A,B) \leq S(A) + S(B) \\ S(A,B) \geq |S(A) + S(B)|. \end{gathered} \tag{17}$$

Equality is only reached when the two systems are not entangled, ie. $\rho_{AB} = \rho_A \otimes \rho_B$. This is in line with classical intuition, as when correlations exist between the two systems the overall uncertainty should be less than if they were totally independent. A stronger property can also be proved, of which normal subadditivity is a

special case. This is known as strong subadditivity. For three systems, A, B and C we have,

$$\begin{aligned} S(A,B,C) + S(B) &\leq S(A,B) + S(B,C) \\ S(A) + S(B) &\leq S(A,C) + S(B,C). \end{aligned} \tag{18}$$

These features of von Neumann entropy have many important implications. For example they put an upper bound on the entanglement between systems. Take a system in which two of the subsystems, A and B, are maximally entangled. If the system is any other state than $\rho_{ABC} = \rho_{AB} \otimes \rho_C$, it can be shown that subadditivity is violated. So if A and B are maximally correlated then they must both be independent of C. This is another difference between quantum and classical information. Classically we can correlate any set of systems as much as we want, where as there is a limit to how much quantum systems can be entangled.

From an operational point of view sharp contrasts can also be found. Consider the classical process in which the state of one system is determined and another system is subsequently put into the same state. There is no obvious reason wy this should not be possible, and operations like this occur constantly in any classical computer. It is therefore suprising that no equivalent procedure is possible for quantum states. For two quantum systems ,A and B, there no unitary operation U, such that for any state in A we have,

$$U(|\psi\rangle_A \otimes |e\rangle_B) = e^{i\alpha} |\psi\rangle_A \otimes |\psi\rangle_B , \tag{19}$$

where α is a real number, and $|e\rangle_B$ is an arbitrary reference state. The proof of the "no-cloning theorem" shows that only states which are parralell or orthogonal can be cloned by the same operator. The theorem puts many limitations on quantum algorithms, like prohibiting the creation of perfect backups for purposes of error correction. However quantum alternatives to classical error correction were soon found.[4]

A similar theorem is known as the "no-teleportation theorem." This states that quantum information cannot be transferred by the exchange of classical information. If we consider the act of physical measurement as being a transfer of quantum information to classical, then the the theorem implies that no possible sequence of measurements will allow us to reconstruct the state in another system.[6] This is strange as there is an obvious method for transferring classical information through a quantum channel, ie. sending a sequence of bits to a system of qubits in orthogonal, and hence distinguishable, states, then measuring each qubit to determine the original input. For example, $(00101) \rightarrow |\uparrow\uparrow\downarrow\uparrow\downarrow\rangle$.

It is however possible to transfer quantum information perfectly using entanglement properties. The simplest example of this is the transfer of a single qubit state. This protocol puts one qubit B, into the original state of another qubit A. The process relies on the use of an intermediate qubit, C. The important points to note about the protocol are that it still needs the exchange of classical information in the form of a measurement outcome, but uses entanglement properties as well. So it is not in contradiction of the no-teleportation theorem. The original state of A is also altered meaning that even though we have recreated the state, we have not cloned it. It should be noted how much at odds the above is with classical information, the replication and transfer of which is theoretically trivial. It is perfectly possible to imagine a channel which takes in a sequence of 1's and 0's and sends them with perfect fidelity. Although practically hard to realise, a channel such as this is hypothetically possible. The same simply isn't true for quantum information.

Despite the conceptual and practical difficulties of manipulating and storing quantum information, there are many scenarios where quantum methods outperform classical ones. To see this, it is useful to look at an explicit example.

GROVER'S ALGORITHM

Consider the problem of finding a particular element in an unsorted list. Classically the most efficent way to solve this is the obvious method of checking each element one by one, and querying if it is the required element. This algorithm has $\mathcal{O}(N)$ time complexity, ie. if a list of length N takes on avergae time t to search, then a list of length $2N$ will take roughly time $2t$ to search. If a classical method could search a list in less than N operations it would have to miss out elements in the list, possibly missing the desired object, and would hence not be suitable.

In 1996 Lev K. Grover published a paper outlining a quantum algorithm for database search with $\mathcal{O}(\sqrt{N})$ time complexity. Although several examples of quantum algorithms which could solve particular problems faster than any possible classical algorithm had already been found, these had mostly been for problems specifically contrsucted to showcase quantum methods, with out much practical value, for the example the DeutschJozsa algorithm.[8] Database searching is ubiquitous in computing, and a quadratic increase in speed could prove very useful. Although the algorithm is somewhat technical, it is possible to give a brief outline.

We are given a list of N objects which we label $0, 1, 2, ..., N-1$. Let x_0 be the label of the object we are searching for. The goal of the procedure is to determine x_0. To do this we use an oracle function f, defined by

$$f(x) = \begin{cases} 1, x = x_0 \\ 0, x \neq x_0. \end{cases} \tag{20}$$

An oracle function is one which we have no internal access to. We don't need to specify the inner workings of f, only it's outputs. Every problem will require the construction of a new f. Next we put the list elements in one-to-one correspondence with orthogonal basis states of an N dimensional Hilbert space. This can be achieved through a system of no more than $log_2(N) + 1$ qubits. Call these states $|0\rangle, |1\rangle, ... |N-1\rangle$. We now define an operator U_f, who's action is defined by,

$$U_f |x\rangle = (-1)^{f(x)} |x\rangle = \begin{cases} -|x\rangle, x = x_0, \\ |x\rangle, x \neq x_0. \end{cases} \quad (21)$$

The final ingredient is what's known as Hadamard gate. This is an operator which sends a specific state to an equiprobable superposition of all states in some orthogonal basis of Hilbert space. For example the Hadmard gate for a single qubit is defined by the matrix,

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (22)$$

in the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$. The Hadamard gate for a larger system of n qubits can be constructed by the recursive definition $H_n = H_{n-1} \otimes H_{n-1}$.

For simplicity's sake I will assume we're working in a Hilbert space which is a tensor product of n qubits, and that $N = 2^n$. First we initialize the system to the state $|\uparrow\uparrow .... \uparrow\rangle$, and then apply H_n. The system is now in an equiprobable superposition of the N orthogonal basis elements, labelled $|s\rangle$. Now we iterate the following step roughly $\mathcal{O}(\sqrt{N})$ times.

1. Apply the operator U_f to the state. This inverts the sign of the state $|x_0\rangle$ and leaves the rest untouched.

2. Apply the operator $U_s = 2|s\rangle\langle s| - I$ to the state.

It can be shown that the overall effect of this step is to increase the amplitude of $|x_0\rangle$ by $\mathcal{O}(\frac{1}{\sqrt{N}})$, hence why it is repeated $\mathcal{O}(\sqrt{N})$ times. After these iterations we then measure the system with respect to the orthonormal basis previously described. With high probability we will find the system in state $|x_0\rangle$, therefore allowing us to determine the value of x_0. Although this is a probabalistic algorithm it can achieve very high accuracy. If the iteration is carried out $[\frac{\pi}{4}\sqrt{N}]$ times for example, then it can be shown that the probaility of success, $p > \frac{N-1}{N}$. [10]

It should be stressed that much of the above has been greatly simplified, but the essence is hopefully clear. This is just one of the many areas in which quantum computing methods can vastly outperform classical ones. To see the extent of the speed up consider an unsorted list of a million objects. Classically it would take five hundred thousand operations on average to find a given element. Using Grover's algorithm it would take $[\frac{\pi}{4}1000] = 786$ operations to find the correct element with a probability of over 0.99999.

CONCLUSION

From the above we can see many parralells between quantum and classical information theory. This is to be expected considering that much of quantum information developed as an attempt to modify existing methods from classical information. The idea of entropy for example is prevalent in both, and in all quantum systems a basis can be chosen for which the Shannon and von Neumann entropy conincide in a sense.13

There are also important differences between the two. Most of these arise due to fundamental features of quantum mechanics such as entanglment and ineluctable indetermincacy. These lead to consequences like the no-cloning and no-teleporatation theories, which lack obvious classical analouges. The occurence of negative conditional von Neumann entropy also marks a departure with classical intuition. Despite the increased conceptual and practical difficulty of quantum information it still holds many tantalising prospects, like the algorithm described above and many others which perform classically impossible tasks.

* Electronic address: s1419697@.ed.ac.uk

[1] "A Mathematical Theory of Communication"-Claude E. Shannon, 1948
[2] "Quantum Information Science"-Seth Loyd, 2009
[3] "Quantum Mechanical Models of Turing Machines That Dissipate No Energy"-Paul Benioff, 1982
[4] "Error correcting codes in quantum theory"-A.M. Steane, 1996
[5] "A Three-Stage Quantum Cryptography Protocol"-Subhash Kak, 2008
[6] "Power, Puzzles and Properties of Entanglment"-J. Gruska, I. Imai,2001
[7] "A fast quantum mechanical algorithm for database search"-Lov K. Grover, 1996
[8] "Rapid solution of problems by quantum computation"-David Deutsch, Richard Jozsa,1992
[9] "Grover's quantum search algorithm"-Samuel J. Lomonaco Jr,2000
[10] http://tph.tuwien.ac.at/ oemer/doc/quprog/node17.html