Quantum Neurocomputation and Signal Processing

Article · October 2000

Source: CiteSeer

CITATIONS
0

READS
23

2 authors:

Hamid R Eghbalnia
University of Cincinnati
96 PUBLICATIONS **1,850** CITATIONS
SEE PROFILE

Amir Assadi
University of Wisconsin–Madison
118 PUBLICATIONS **667** CITATIONS
SEE PROFILE

Quantum Neurocomputation and Signal Processing

Hamid Eghbalnia, Amir Assadi
University of Wisconsin, Madison

ABSTRACT

In this paper we consider a Quantum computational algorithm that can be used to determine (probabilistically) how close a given signal is to one of a set of previously observed signal stored in the state of a quantum neurocomputional machine. The realization of a new quantum algorithm for factorization of integers by Shor and its implication to cryptography has created a rapidly growing field of investigation. Although no physical realization of quantum computer is available, a number of software systems simulating a quantum computation process exist. In light of the rapidly increasing power of desktop computers and their ability to carry out these simulations, it is worthwhile to investigate possible advantages as well as realizations of quantum algorithms in signal processing applications. The algorithm presented in this paper offers a glimpse of the potentials of this approach. Neural Networks (NN) provide a natural paradigm for parallel and distributed processing of a wide class of signals. Neural Networks within the context of classical computation have been used for approximation and classification tasks with some success. In this paper we propose a model for Quantum Neurocomputation (QN) and explore some of its properties and potential applications to signal processing in an information-theoretic context. A Quantum Computer can evolve a coherent superposition of many possible input states, to an output state through a series of unitary transformations that simultaneously affect each element of the superposition. This construction generates a massively parallel data processing system existing within a single piece of hardware. Our model of QN consists of a set of Quantum Neurons and Quantum interconnections. Quantum neurons represent a normalized element of the n-dimensional Hilbert space - a state of a finite dimensional quantum mechanical system. Quantum connections provide a realization of probability distribution over the set of state that combined with the Quantum Neurons provide a density matrix representation of the system. A second layer with a similar architecture interrogates the system through a series of random state descriptions to obtain an average state description. We discuss the application of this paradigm to the quantum analog of independent states using the quantum version of the Kullback-Leibler distance.

INTRODUCTION

Quantum computation spurred recently by the work of Shor, has led to a new consideration of information theoretic notions which can be used to study these questions. It has been suggested that the modern digital computer inherited its

atomistic mode of computation from the concepts proposed by the Neuron Doctrine following the pioneering work of Ramon y Cajal. Cross-fertilization from the fields of synthetic computation and natural computation has led to ideas such as parallel and distributed processing, soft computing, associative memory, as well as application of information theoretical techniques in computing to evaluation of information carrying capacity of spikes in the brain. However, evidence to date seems to point to the unavoidable fact that the very problems that are most difficult for computers to address are problems that are routinely handled by natural computation and those that are routinely handled by computers pose the most difficulty for natural computing systems. For example, extraction of spatial structure (pattern recognition) from a visual scene in the presence of occlusions, perspective distortions, intensity variations and shadows are effortless tasks for the visual system. While perfect recall of vast quantities of data, indefinite storage, perfect repetition and similar tasks continue to be in the domain of synthetic computation.

In light of shortcomings of current computational model in certain applications, it seems appropriate to consider alternative information-theoretic measures based on alternative computational paradigms that could be used to analyze and possibly expand our current arsenal of algorithms. We emphasize that our work is to investigate alternative computational models and study how the solution of given problems can be cast (and possibly simplified) in the new setting. Although quantum computers do not exist, simulation of algorithms suggested by this paradigm on standard machines may offer practical alternatives for specific problems. A discussion of the ideas can be found in [1,2,3,4,6,8] and references therein. In this paper we focus on some early work in formulating a computational model for distinguishing states of a system. The outline of the paper is as follows. After a brief discussion of our model in this section we then conduct a brief review of quantum information theory and establish the notation in this paper. We then follow with examples that illustrates how new possibilities arise in the quantum computational paradigm. We then describe the quantum analog of kullback-Leibler distance and describe an algorithms for using this measure to distinguish states.

Motivated by the channel theory in the human visual system, we consider an information theoretic model based on the principals of quantum computation. We consider encoding of signals in a quantum system and investigate algorithms for retrieval of such information. A simple visual aid for understanding the quantum system can be realized as a system composed of a two fully interconnected layers of quantum neurons as depicted in figure 1.

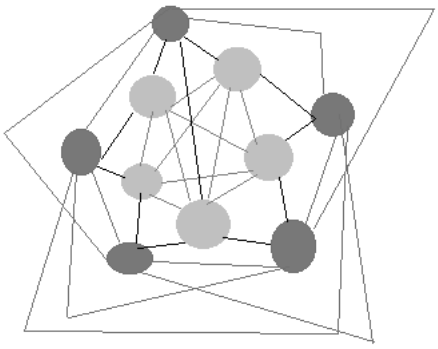

Figure 1. Layer A is the depicted in lighter (gray) color and layer B is depicted in the darker (Green) color with corresponding colors for inter-layer and black for intra-layer connections. Layer A maintains the state of the system and layer B will be used to prepare random states for measuring the state of the system.

Roughly speaking, layer A maintains the state of the system and layer B is used to ascertain an average property which can be used to distinguish states.

Inherent in quantum computation is the need to address the measurement problem. Five minimal requirements are often cited as necessary for the existence of a quantum computer. A particularly important element is the necessity that it be possible to subject the quantum system to "strong" measurements. "Strong" simply refers to measurement that determines which orthogonal eigenstate of some particular Hermitian operator the quantum state belongs to. This "Strong" measurement will, at the same time, project the wavefunction of the system irreversibly into the corresponding eigenfunction. One standard example of such a measurement is the Stern-Gerlach experiment in which the z-component of a spin-1/2 particle is projected into one of its two eigenvalues. Unfortunately, many actual quantum measurements in the laboratory are of a "weak" nature. For a complete discussion we refer the reader to the textbooks [5,7,9] , and outline the idea here for our purposes.

The individual quantum system, say a single spin-1/2 system, might interact very weakly with the measurement apparatus. This means: the probability that the apparatus registers "spin up" is only weakly correlated with the actual wavefunction amplitude for the spin to be up. A more quantitative statement would say that there exist weak measurements on a state $a \mid \uparrow\rangle + b \mid \downarrow\rangle$ such that one can obtain "spin up" with probability

$$P_{\uparrow} = \delta\left(|a|^2 - 1/2\right) + 1/2$$

for arbitrarily small δ . After such measurements the quantum state of the system has been disturbed hardly at all - thus the term weak. However, in "weak" measurements little information has been gained by the measurement about the state of the spin. In some cases weak measurements are satisfactory for learning about the quantum properties of systems. Examples of these cases occur when they can be done on macroscopically large ensembles, involving either many replicas of the same quantum system (e.g. typical in NMR), or when many identical runs of the same quantum measurement can be performed. By averaging over such ensembles, good knowledge of the state a above can be obtained, no matter how small δ is. It is not clear whether these weak measurements satisfy the requirements in quantum computation. However, this is a fundamental problem where progress appears to be taking place.

QUANTUM INFORMATION THEORY

In this section we briefly review aspect of the quantum formalism and quantum information theory needed for our discussion. In the quantum formalism the state of a system with a finite set of states T is characterized by a unit vector (up to a phase factor of $e^{i\phi}$) $|\psi\rangle$ in the complex Hilbert space $\mathcal{H}$ that is equipped with a Hermitian scalar product $\langle .|.\rangle$. Corresponding to each classical states a we have the standard vectors $|a\rangle \in \mathcal{H}$ which form an orthonormal basis of $\mathcal{H}$. Time evolution of a quantum system is given by a transformation via a unitary operator: $\psi\rangle = U|\psi\rangle$ Where U represents the unitary operator. Elements of $\mathcal{H}$ are usually denoted as $|\zeta\rangle$, and the scalar product is denoted by $\langle\zeta|\eta\rangle$. Therefore, it is natural to think of $\langle\zeta|$ as defining a linear functional $\langle\zeta|:|\eta\rangle \mapsto \langle\zeta|\eta\rangle$ on $\mathcal{H}$ and the space of such functionals is denotes by $\mathcal{H}^*$. Elements of $\mathcal{H}$ and $\mathcal{H}^*$ have a coordinate representation and in this representation one can think of them as column and row vectors respectively. Based on the above, given two vectors, a natural construction for a linear operator on $\mathcal{H}$ can be given as:

$$|\zeta\rangle, |\eta\rangle \in \mathcal{H} \Rightarrow |\eta\rangle\langle\zeta| \in \mathcal{H} \otimes \mathcal{H}^*$$

We denote the result of application of a linear operator $\mathrm{A}: \mathcal{H} \to \mathcal{H}$ to a vector as $\mathrm{A}|\zeta\rangle$. The following notation is common and useful:

$$\langle\zeta|\mathrm{A}|\eta\rangle = \langle\zeta|\mathrm{A}\,\eta\rangle = \langle\mathrm{A}^{\bullet}\zeta|\eta\rangle$$

where $\mathrm{A}^{\bullet}$ is the adjoint.

Orthogonal projection of a vector $|\zeta\rangle$, denoted by $\Pi|\zeta\rangle$, is projection onto a linear subspace of $\mathcal{H}$ This projection operator is used in giving the probabilistic

interpretation of the quantum formalism as well as the relation between the system and its subsystems. In an analogous manner to standard probability theory, the quantum probability depends on a quantum state and a linear subspace. From these two, the probabilistic interpretation is given as:

$$P = \langle \zeta | \Pi | \zeta \rangle = Tr(\rho \Pi)$$

Where $\rho = |\zeta\rangle\langle\zeta|$ is called the density operator associated to the state $|\zeta\rangle$. A generic density operator is said to represent a *mixed state* of the system, while quantum states defined so far are called *pure*. Density operators evolve through unitary operations as well according to: $\rho \mapsto U \rho U^{\bullet}$.

The concept of information in quantum physics is introduced through Von Neumann entropy as:

$$S(\rho) = -Tr \rho \log \rho$$

where ρ represents the mixture state defined by $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ where p_i is the probability of occurrence of the state $|\psi_i\rangle$. The relative entropy of a pair of states is defined as:

$$\mathcal{D}(\rho \| \eta) = S(\rho) - Tr \rho \log \eta$$

and defines a directed distance between the two states which enjoys most properties of a distance function except symmetry. Similar to the results in classical probability theory the probability for erroneous identification of a state after N measurement decreases as an exponential with exponent of N times the distance between the two states to be measured. Corresponding to the mutual information in classical entropy and the bound it establishes on a communication channel, Holevo established the following upper bound on the mutual information as:

$$\chi = S(\rho) - \sum_i p_i S(\rho_i) = \sum_i p_i \mathcal{D}(\rho_i \| \rho) = \sum_i p_i \mathcal{D}(\rho_i \| \eta) - \mathcal{D}(\rho \| \eta)$$

This quantity expresses an average relative entropy from the average state ρ to the members of the signal ensemble.

MOTIVATING EXAMPLES

The following examples serve two purposes: 1) they illustrates some of the notation introduced above and 2) they demonstrates new coding possibilities in quantum states as well as nonlinear behavior of the sum of two interfering observables. These examples motivates our interest in quantum algorithms.

Example1: Independent States: Suppose we are given X and Y as independent states both having densities that are half |0> and half |1>. Then, this computation illustrates that we learn nothing about x by measuring y (as expected).

$$\rho_x = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \qquad \rho = \rho_x \otimes \rho_y = \begin{pmatrix} 1/4 & & & \\ & 1/4 & & \\ & & 1/4 & \\ & & & 1/4 \end{pmatrix}$$

$$S(x) = S(y) = 2(-\frac{1}{2}\ln(\frac{1}{2})) = 1$$

$$S(x \otimes y) = 4\left(-\frac{1}{4}\ln\left(\frac{1}{4}\right)\right) = 2$$

$$S(x \mid y) = S(x \otimes y) - S(y) = 1$$

$$I(x, y) = S(x) + S(y) - S(x \otimes y) = 0$$

Example2:Classically Correlated States. This example illustrates the case for completely correlated states. Here x and y are classically correlated, (either both |0> or both |1>), and that |00) and |11) are equally likely. Then, we have:

$$\rho = \begin{pmatrix} 1/2 & & & \\ & 0 & & \\ & & 0 & \\ & & & 1/2 \end{pmatrix}$$

$$S(x) = S(y) = 2(-\frac{1}{2}\ln(\frac{1}{2})) = 1$$

$$S(x \otimes y) = 4\left(-\frac{1}{4}\ln\left(\frac{1}{4}\right)\right) = 1$$

$$S(x \mid y) = S(x \otimes y) - S(y) = 0$$

$$I(x, y) = S(x) + S(y) - S(x \otimes y) = 1$$

Example3: Quantum Entangled States. In this example x and y are in the entangled state |00> + |11> (normalized). The combined state is a pure state that has zero entropy while a measurement of x or y separately will produce |0> and |1> with equal probability. This produces the unexpected result that the mutual information is greater than the information in either one of the subsystems. We also notice that the conditional entropy is negative. The extra information in I(X,Y) is sometime referred to as *virtual* information (cannot convey information).

$$\rho = \begin{pmatrix} 1/2 & & & 1/2 \\ & 0 & & \\ & & 0 & \\ 1/2 & & & 1/2 \end{pmatrix}$$

$$S(x \otimes y) = -1\ln 1 - 3(0\ln 0) = 0$$

$$S(x \mid y) = S(x \otimes y) - S(y) = -1$$

$$I(x, y) = S(x) + S(y) - S(x \otimes y) = 2$$

Example4: Consider the transmission of sequence of qubit signal states of $\psi_1 = |0\rangle$ and $\psi_2 = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ where each transmission can be one of the states with equal probability of 0.5. The density matrix of the source is: $\rho = 0.5|\psi_1\rangle\langle\psi_1| + 0.5|\psi_2\rangle\langle\psi_2|$ and the Von Neuman entropy can be computes as defined above. The communication of this message can be clearly accomplished by transmitting one qubit per state. However, quantum source coding theorem shows that the signal can be asymptotically transmitted using S qubits per state where S is the Von Neuman entropy. In the above example a direct calculation yields S=0.601 qubits.

Example5: Consider the Hilbert space $\mathcal{H} = \mathbb{C}^2$ with the canonical basis $e_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $e_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then, the space of observable on Hilbert space, $\mathcal{O}(\mathcal{H})$, has as basis the well-known Pauli spin matrices:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

This is an orthonormal basis for $\mathcal{O}(\mathcal{H})$ with the scalar product $\langle X, Y \rangle = \frac{1}{2} Tr\, XY$. In the pure state e_0 the observables σ_1, σ_2 and σ_3 assume the values 1, -1 and 1 respectively with probabilities 0.5,0.5 and 1. However, an elementary calculation shows that the observable $\sigma_1 + \sigma_2$ has eigenvalues $\pm\sqrt{2}$, whereas each observable takes has eigenvalues ± 1 individually. This points out the fact that the vectorial sum of two observables in $\mathcal{O}(\mathcal{H})$ has an essentially nonlinear feature when the observables are interfering observables. This stands in sharp contrast to the way sums of random variables behave in classical probability.

INDEPENDENT COMPONENT ANALYSIS

In classical probability theory mutual information is a measure of channel capacity. Mutual information $I(X;Y) = H(X) + H(Y) - H(X,Y)$ is defined in terms of Shannon information $H(X) = -\sum_x p_x \log p_x$ of an ensemble $\{X, p_X\}$ and measures the number of bits of information obtained about X by observing Y. The Kullback-Leibler (KL) divergence between the distributions p and q is defined as $K(p \mid q) = \int p \ln(p/q)$ (the integral is replaced by a sum for discrete distributions). The KL distance can be viewed as the mean information required to turn the prior q into the posterior p. Thus, Mutual information and KL distance both measure relative information between two distributions. This information theoretic point of view has motivated a number of applications in machine vision as well as human visual system analysis. Consider a finite sample $(x_1, \cdots, x_n)$ of d-vectors representing a set of measurements of the "state of nature". The key computational ingredient of Independent Components Analysis (ICA) is to find a linear map that transforms the observed multi-variate data into a new collection of statistically independent components. There are a number of different, mostly equivalent, ways to formulate Independent Components Analysis. Most of these methods amount to minimizing the Kullback-Leibler divergence between the joint probability and marginal probabilities of the output signals. As is clear from this discussion, ICA is a linear approach to separating mixtures. A number of approaches for extending the ICA approach to include nonlinearities has been suggested. However, as noted in the example in the preceding section, nonlinear affects occur naturally as a byproduct of the quantum physical approach to states of information and is motivated by success of experimental quantum physics.

QUANTUM COMPONENT ANALYSIS

Consider a system of n neurons each capable of representing an m-qubit state and a set of probability amplitudes for the states. This can be represented as an ensemble of signals $\{|\psi_1\rangle, \cdots, |\psi_n\rangle; p_1, \cdots, p_n\}$ with the density matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. We are given an arbitrary signal (a state) and we wish to determine if this signal is close to, or can be approximated by, one of the original states used to construct the density matrix. The information available to us is the density matrix (the entangled states) and not the individual states. Therefore, we would like to characterize the mean information required to distinguish the new

state. Given a signal η, the KL distance for a quantum system is a good measure of the mean information and is defined as:

$$K(\rho\,|\,\eta) = \sum_j Tr\left(\rho E_j\right)\log\left(Tr\left(\rho E_j\right)/Tr\left(\eta E_j\right)\right)$$

Where *E* is a positive operator valued measure which describes the basis in which the state is measured. A bad choice of basis can make states less distinguishable. Therefore, one must maximize *K* above over all possible basis in order to obtain an optimal KL distance. The optimization over the space E seriously limits the utility of this formula. We note here that Holevo bound requires that we maximize the average distance from the average state to the individual members of the signal ensemble in order to achieve maximum distinguishability of states. Thus, the original encoding scheme of the states plays an important role in determining distinguishability via the KL distance.

Before we proceed to show how to estimate the mean information, we first show how this can be equivalently represented in the quantum neurocomputational model we have proposed. Consider the Gram matrix $G_{ij} = \sqrt{p_i p_j}\left\langle \psi_i \middle| \psi_j \right\rangle$. This matrix has the following properties: (1) its nonzero eigenvalues coincide with those of the density matrix above and thus it has the same entropy, (2) it is always positive with trace equal to one. In this context, the interconnections of our Quantum neurocomputer represent the square root of the product of probabilities and the states of each quantum neuron are as before. Given this equivalence we can continue to formulate our algorithm in the context of density matrices realizing that a direct translation to the model above is clear.

Since the KL distance is formulated in terms of the trace operator we wish to have an effective method of estimating trace. The estimation of trace of a matrix M in a d-dimensional Hilbert space spanned by a set of orthonormal states is simple to compute. However, the orthonormal states in which the trace must be computed is not known to us since as discussed before the KL distance must be optimized over the basis E. We show here that this computation can be approximated using sampling over a set of random states using our quantum computer's ability to efficiently and rapidly compute observations of the form $P = \left\langle \zeta \middle| \Pi \middle| \zeta \right\rangle = Tr(\rho\Pi)$.

Consider any set of orthonormal states $\left\{\left| e_n \right\rangle\right\}$ and d complex numbers, drawn randomly (with respect to each component of the complex number) form elements on the unit circle (so that they have zero mean) designated by a_i. Form the random vector $\left|\zeta\right\rangle = \sum_{n=1}^{d} a_n \left| e_n \right\rangle$ and use this vector to compute the trace of the matrix M:

$\langle \zeta | M\zeta \rangle = \sum_{i=1}^{d}\sum_{j=1}^{d} a_i^* a_j \langle e_i | Me_j \rangle$. Consider R independent realization of the random vectors above and examine the average value. $R^{-1}\sum_{k=1}^{R}\sum_{j=1}^{d}\sum_{i=1}^{d} a_{ik}^* a_{jk} \langle e_i | Me_j \rangle$ and note that based on our assumptions on the distribution of random complex numbers a, we have: $\lim_{R\to\infty} R^{-1}\sum_{k=1}^{R} a_{ik}^* a_{jk} = \mathbb{I}$. It follows immediately that the trace can be approximated by averaging over a set of random states. Therefore, this method provides a computational approach to the estimation of the Quantum KL distance in a quantum computer modeled by our quantum neurocomputer. This model offers the opportunity of simulations based on software systems meant to simulate quantum computation. Given a system specified by an initial set of complex state vectors (such as our quantum neurocomputer) in a superposition of states, we have devised an algorithm for determining how much information is contained in a given state about our system. This method may open new avenues for signal understanding.

REFERENCES

[1] Adleman L, Demarrais J and Huang M-D, Quantum Computability, SIAM Journal of Computation 26 5 pp 1524-1540 October, 1997

[2] Bennett C H, Brassard G, Cr'epeau C, Jozsa R, Peres A and Wootters W K Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. 70 1895- 1898, 1993

[3] Bennett C H, Bernstein E, Brassard G and Vazirani U Strengths and Weaknesses of quantum computing, SIAM Journal of Computation 26 5 pp 1510-1523 October, 1997

[4] Buhrman H, Cleve R and Wigderson A, Quantum vs. Classical Communication and Computation, in Proc. of the 30th Annual ACM Symposium on Theory of Computing (STOC) (1998)

[5] Cirac J I, Zoller P, Kimble H J and Mabuchi H Quantum state transfer and entanglement distribution among distant nodes of a quantum network, Phys. Rev. Lett. 78, 3221, 1997

[6] Deutsch D and Jozsa R, Rapid solution of problems by quantum computation, In Proc. Roy. Soc. Lond A 439 553-558, 1992.

[7] Peres, A. (1993) Quantum theory: concepts and methods, Kluwer Academic Publishers, Dordrecht.

[8] Simon D 1994 On the power of quantum computation, SIAM J. Comp., 26, No. 5, pp 1474-1483,October 1997.

[9] Wheeler J A and Zurek W H, eds, 1983 Quantum theory and measurement (Princeton Univ. Press,Princeton, NJ).