

Pattern Classification Using a Quantum System

Dan Ventura

Brigham Young University
Department of Computer Science
Provo, UT 84602 USA
ventura@cs.byu.edu
<http://axon.cs.byu.edu/Dan>

We consider and compare three approaches to quantum pattern classification, presenting empirical results from simulations.

Introduction

We consider the possibility of using a quantum system to classify binary patterns. Motivations for doing so include the potential for classification systems to exist on a scale not possible with current classical logic circuits and the potential for quantum systems to perform computations not feasible on classical systems [1] [2] [3] [4].

The idea of using quantum systems as neural information processing systems is not new [5] [6] [7] [8] [9] but is not yet well understood. Here, we present some analysis and empirical results for furthering such investigations.

Specifically, let $\mathbf{B}=\{0,1\}$ and let $T=\{(x_i, y_i)\}$ be a set of m pairs of points x_i in \mathbf{B}^n and labels y_i in \mathbf{B} . We would like to construct a quantum system for correctly labeling those points in T and for generalizing in a reasonable way to label other points $z \in \mathbf{B}^n$ with $z \notin T$. In other words, we would like to construct a quantum classification system that approximates the function $f: \mathbf{B}^n \rightarrow \mathbf{B}$ from which the set T was drawn. The use of entanglement for just such a task has been discussed in [10]. We consider three variations on incorporating the information contained in the set T into a quantum classification system based on Grover's iterative search algorithm [2]. The three methods will be described and compared and empirical simulation results from some very simple classification problems will be presented.

Approach

Quantum computation is based upon physical principles from the theory of quantum mechanics, which is in many ways counterintuitive. Yet it has provided us with perhaps the most accurate physical theory (in terms of predicting experimental results) ever devised by science. The theory is well established and is covered in its basic form by many

textbooks (see for example [11]). Quantum computation is a relatively new discipline and not yet completely understood; however, [12] provides an excellent introduction to many of the key ideas.

A quantum system is described by a superposition of a set of basis states, $|\phi_i\rangle$, that span a Hilbert space:

$$|\psi\rangle = \sum_i c_i |\phi_i\rangle \quad (1)$$

$|\psi\rangle$ is said to be in a superposition of the basis states $|\phi_i\rangle$, and the coefficients c_i may be complex with

$$\sum |c_i|^2 = 1 \quad (2)$$

Use is made here of the Dirac bracket notation, where the ket $|\cdot\rangle$ is analogous to a column vector, and the bra $\langle\cdot|$ is analogous to the complex conjugate transpose of the ket.

Quantum computation consists of state preparation, effecting useful time evolution of the quantum system, and measurement of the system to obtain information. Upon measurement, the system will probabilistically "collapse" to a single basis state, and the object of quantum computation is to attempt to ensure that the measured basis state is with high probability one which gives the correct answer or desired information.

Here we consider three different approaches to state preparation based on the information in the set T . Classification of instances is performed using Grover's search algorithm [2] (the time evolution step) and final observation of the system. In what follows we assume a set of $n+1$ two-state quantum systems, and for convenience we will label the two states $|0\rangle$ and $|1\rangle$.

Learning the Set T

We will consider three different methods for learning the pattern classifications by representing them in a quantum superposition: *inclusion*, *exclusion*, and *phase inversion*. Inclusion is perhaps the most intuitive and is the approach suggested in [10]. It represents each of the labeled points in T as a basis

state in the superposition with a nonzero coefficient. Basis states corresponding to points not in T have zero coefficients:

$$|\psi\rangle = \frac{1}{\sqrt{m}} \sum_{(x_i, y_i) \in T} |x_i y_i\rangle \quad (3)$$

Exclusion is the opposite approach, including each point *not* in T with a nonzero coefficient while those points in T have zero coefficients. This approach is employed in [13] in implementing a quantum associative memory.

$$|\psi\rangle = \frac{1}{\sqrt{2^n - m}} \sum_{(x_i, y_i) \notin T} |x_i y_i\rangle \quad (4)$$

Phase inversion includes all basis states in the superposition with coefficients of equal amplitude but with differing phases based on membership in T :

$$|\psi\rangle = \frac{1}{\sqrt{2^n}} \left(\sum_{(x_i, y_i) \notin T} |x_i y_i\rangle - \sum_{(x_i, y_i) \in T} |x_i y_i\rangle \right) \quad (5)$$

As an example, consider the case of $n=2$ and $T=\{(00,0), (11,1)\}$. The inclusive method using Eq. 3 produces the superposition

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|000\rangle + |111\rangle)$$

The exclusive method using Eq. 4 produces the superposition

$$|\psi\rangle = \frac{1}{\sqrt{6}} (|001\rangle + |010\rangle + |011\rangle + |100\rangle + |101\rangle + |110\rangle)$$

And the phase inversion method using Eq. 5 produces the superposition

$$|\psi\rangle = \frac{1}{\sqrt{8}} (-|000\rangle + |001\rangle + |010\rangle + |011\rangle + |100\rangle + |101\rangle + |110\rangle - |111\rangle)$$

Classification

Since the intent here is to focus on the methods for representing the set T in a quantum system, we will use a straight forward approach to classification, employing Grover's iterate. This can be described as a product of unitary operators GR applied to the quantum state iteratively. This produces a periodic behavior that can be predicted, and the probability of the desired result maximized by measuring the system after an appropriate number of iterations. The operator R is a

phase inversion of the state(s) that we wish to observe upon measuring the system and is represented by the identity matrix I with those diagonal entries corresponding to the desired state(s) equal to -1 . The operator G has been described as an inversion about average and if $|\psi\rangle = \frac{1}{\sqrt{2^{n+1}}} \sum_{(x_i, y_i) \in \mathbf{B}^{n+1}} |x_i y_i\rangle$, then G can be represented as the matrix $2|\psi\rangle\langle\psi| - I$.

Continuing the example, if we wished to classify the point 00,

$$R = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$G = \frac{1}{4} \begin{bmatrix} -3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -3 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -3 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -3 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -3 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -3 \end{bmatrix}$$

Figure 1 shows the probability of observing the correct classification upon system measurement vs. the number of iterations of Grover's search applied to the exclusive superposition obtained using the example set T and Eq. 4. The solid line represents the probability P_c of a correct classification (0), and the dotted line the probability P_w of an incorrect classification (1). Note that these two probabilities do not sum to one. This is because there is a nonzero probability $P_r = 1 - (P_c + P_w)$ of an irrelevant classification (classifying a pattern other than the one in which we are interested). The periodic nature of the algorithm is clearly discernable, and it can be seen that in this case the probability of success is maximized after four iterations. We will be interested not only in the maximum of P_c but also more particularly in the ratio P_c/P_w . This is because it is easy upon measurement to determine whether or not we have made an irrelevant classification (in which case we can simply perform the classification again); however, if we have classified the desired pattern, it is impossible to know for sure whether we have made a

correct classification. For this reason we are interested in large values for the ratio P_c/P_w – the larger the ratio, the higher the confidence that the classification is correct.

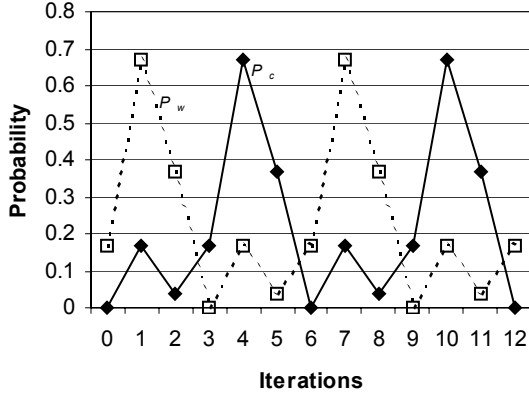


Figure 1. Probability of correct classification for exclusive representation of a set of two example patterns.

For example, Figure 1 shows that after four iterations the probability of classifying the desired pattern (00) is $P_c + P_w \approx 0.67 + 0.17 = 0.84$. Therefore there is a probability $P_r = 0.16$ that a pattern other than 00 will be classified. If the desired pattern is classified, there is a probability $P(P_c | \sim P_r) = P_c / (P_c + P_w) \approx 0.8$ that the classification will be correct. The higher the ratio P_c/P_w , the higher the conditional probability.

Figure 2 compares the three methods for encoding patterns in a quantum system with regards to the probability of an irrelevant classification. Note that the exclusion method demonstrates the lowest overall probability of irrelevant classification.

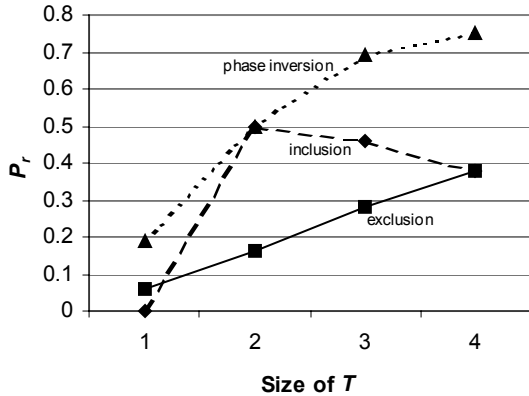


Figure 2. Probability of irrelevant classification for inclusion, exclusion and phase inversion.

Since it is obvious when an irrelevant classification is made, this probability simply governs the number of times the classification will have to be repeated. On average this will be $1/(1 - P_r)$.

Figure 3 makes a similar comparison amongst the three methods for the conditional probability of a correct classification, given that the classification is not irrelevant. Here the inclusive method appears to be the best choice. Recall that this probability is more critical than the one shown in Figure 2 because there is no way of knowing whether the classification is correct or not; thus a high conditional probability is desirable.

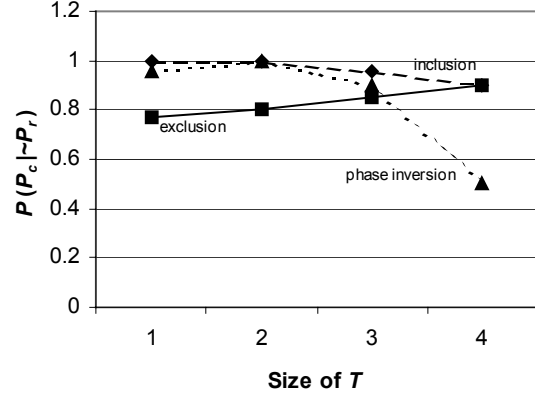


Figure 3. Conditional probability of correct classification (given that the classification is not irrelevant).

Generalization

To this point nothing has been done to facilitate generalization, and new patterns will be classified randomly. This can be changed by modifying Eqs. 3-5 so that states whose labels are close (according to some metric) will have similar coefficients. Therefore, known patterns from the set T will influence the classification of similar unknown patterns.

For the inclusion method, this may be accomplished by measuring Hamming distance from a known pattern and setting the coefficient proportionally. If the pattern (x_i, y_i) is in the set T , then for patterns (x_j, y_j) not in T , the coefficient

$$c_j = \frac{\sqrt{h(x_i, x_j) + s}}{r}, \text{ and the coefficient } c_i = \frac{\sqrt{n + t}}{r},$$

where $h(x_i, x_j)$ gives the Hamming distance and r, s and t are chosen to appropriately weight the patterns and to maintain the unitarity demanded by Eq. 2. For the exclusion method, a similar approach sets the

$$\text{coefficients as } c_j = \frac{\sqrt{n - h(x_i, x_j) + s}}{r} \text{ and } c_i = 0.$$

For the phase inversion method, information is stored in the relative phases of the states, all of whose coefficients have the same magnitude. Since the coefficients can be complex, we can rotate patterns “partially” out of phase to facilitate generalization,

$$c_j = \frac{e^{ih(x_i, x_j)/n}}{\sqrt{2^n}} \text{ and } c_i = \frac{-1}{\sqrt{2^n}} \text{ where we slightly}$$

abuse notation by using i both as a subscript and to represent $\sqrt{-1}$.

Figures 4 and 5 show results for incorporating such generalization into a system learning the very simple set $T=\{(00,0)\}$. As before, exclusion appears to produce the lowest probability of irrelevant classification; however, inclusion produces better conditional probability of correct classification. Interestingly, at least in this simple example, performance for unknown patterns is actually better than that for known patterns.

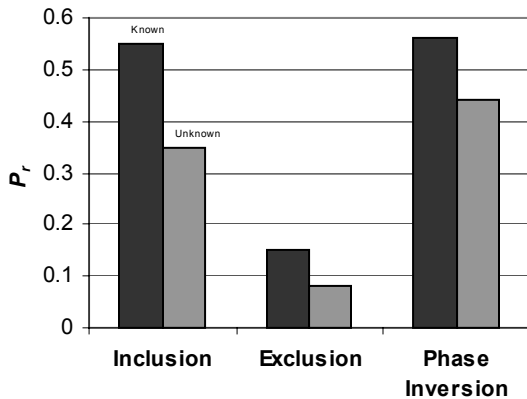


Figure 4. Probability of irrelevant classification for generalizing system for both known and unknown patterns.

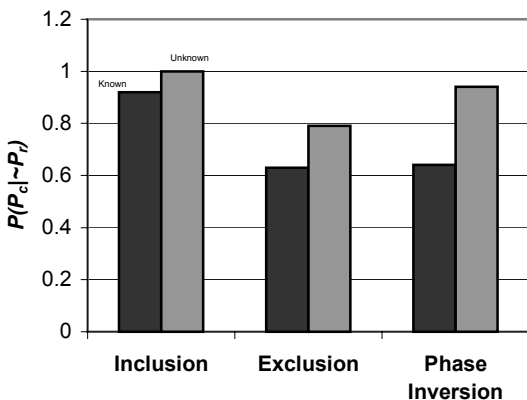


Figure 5. Conditional probability of correct classification (given that the classification is not irrelevant) for generalizing system for both known and unknown patterns.

Discussion

We have presented empirical results comparing three approaches to quantum pattern classification. These results favor the inclusion method based on the fact that the inclusion method appears to produce the best conditional probability of correct classification. It should be emphasized, however, that these results are

not yet conclusive and that they apply only to the case where Grover's search is involved – other less general approaches may be more useful in pattern classification. If probability of irrelevant classification is considered, the exclusion method appears to exhibit the best performance, and it may be interesting to consider an inclusive/exclusive hybrid approach.

Finally, we note that an efficient method for constructing the state shown in Eq. 3 is presented in [14], and the state shown in Eq. 5 can easily be generated using simple rotation matrices; the state shown in Eq. 4 can perhaps be generated efficiently using something akin to the reverse of the algorithm in [14], but to the author's knowledge, this has not yet been shown conclusively.

References

- [1] P. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer", *SIAM Journal of Computing*, **26**(5), 1997, pp. 1484-1509.
- [2] L. Grover, "A Fast Quantum Mechanical Algorithm for Database Search", *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing*, ACM, New York, 1996, pp. 212-219.
- [3] D. Simon, "On the Power of Quantum Computation", *SIAM Journal of Computing*, **26**(5), 1997, pp. 1474-1483.
- [4] D. Deutsch and R. Jozsa, "Rapid Solution of Problems by Quantum Computation", *Proceedings of the Royal Society of London Series A*, **439**, 1992, pp. 553-558.
- [5] M. Perus, "Neuro-quantum Parallelism in Brain-mind and Computers", *Informatica* **20**, 1996, pp. 173-183.
- [6] A. Narayanan and T. Meneer, "Quantum Artificial Neural Network Architectures and Components", *Information Sciences* **128**, 2000, pp. 231-255.
- [7] M. Zak, "Quantum Decision-maker", *Information Sciences* **128**, 2000, pp. 199-215.
- [8] E. Behrman, L. Nash, J. Steck, V. Chandrashekar and S. Skinner, "Simulations of Quantum Neural Networks", *Information Sciences* **128**, 2000, pp. 257-269.
- [9] D. Ventura and T. Martinez, "Quantum Associative Memory", *Information Sciences* **124**, 2000, pp. 273-296.
- [10] D. Ventura, "On the Utility of Entanglement in Quantum Neural Computing", *Proceedings of the International Joint Conference on Neural Networks*, 2001, pp. 1565-1570.
- [11] R. Feynman, R. Leighton and M. Sands, *The Feynman Lectures on Physics*, Addison-Wesley Publishing Company, Massachusetts, 1965.
- [12] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.
- [13] A. Ezhov, A. Nifanova and D. Ventura, "Distributed Queries for Quantum Associative Memory", *Information Sciences* **128**, 2000, pp. 271-293.
- [14] D. Ventura and T. Martinez, "Initializing the Amplitude Distribution of a Quantum State", *Foundations of Physics Letters*, **12**(6), 1999, pp. 547-559.