

Classical Artificial Neural Network Training Using Quantum Walks as a Search Procedure

Luciano S. de Souza^{ID}, Jonathan H. A. de Carvalho^{ID}, and Tiago A. E. Ferreira^{ID}

Abstract—This article proposes a computational procedure that applies a quantum algorithm to train classical artificial neural networks. The goal of the procedure is to apply quantum walk as a search algorithm in a complete graph to find all synaptic weights of a classical artificial neural network. Each vertex of this complete graph represents a possible synaptic weight set in the w -dimensional search space, where w is the number of weights of the neural network. To know the number of iterations required *a priori* to obtain the solutions is one of the main advantages of the procedure. Another advantage is that the proposed method does not stagnate in local minimums. Thus, it is possible to use the quantum walk search procedure as an alternative to the backpropagation algorithm. The proposed method was employed for a *XOR* problem to prove the proposed concept. To solve this problem, the proposed method trained a classical artificial neural network with nine weights. However, the procedure can find solutions for any number of dimensions. The results achieved demonstrate the viability of the proposal, contributing to machine learning and quantum computing researches.

Index Terms—Artificial neural networks training, quantum computing, quantum walk, search algorithm

1 INTRODUCTION

THE need for increased computing power and the miniaturization of components at scales where quantum effects cannot be ignored [1] support the idea that quantum physics can redefine a new frontier for computing problems by taking an essential role in the computational race [2]. Quantum effects can provide computational gains and algorithms that are mostly more efficient than their classical counterparts for many problems. Quantum computing seeks through the use of aspects of quantum mechanics to expand computational horizons [3], [4]. An example of the computational power of quantum computing is performing a search for elements in a disordered database in just $O(\sqrt{N})$ [5], where N is the size of the database. The gain is quadratically faster than its classical analog.

In this perspective, efforts are employed in the search for more efficient algorithms. In particular, there is a branch of research called quantum machine learning (QML) [6], where machine learning methods and artificial intelligence are integrated into the world of quantum computing in the hope to find more efficient procedures than classical algorithms. According to V. Dunjko, J. M. Taylor, and H. J. Briegel [7], the quantum processing of information from quantum machine learning is interactively configured in two parts called the agent-environment. These parts are grouped into four categories: *CC*, *CQ*, *QC*, and *QQ*, where *C* means *Classical*, *Q* means *Quantum*. The first letter is referent to the agent

and the second to the environment. *CC* encompasses classical machine learning. *CQ* analyzes how classical learning techniques can help in quantum tasks. *QC* represents the quantum variants of classical learning algorithms facing a classical environment, where this work is classified. *QQ* is the quantum systems world.

Papers grouped within the *QC* category are developed to improve classical procedures with the quantum information processing (QIP) paradigm. E. Aïmeur, G. Brassard, and S. Gambs [8] showed that it is possible to accelerate unsupervised learning algorithms by quantizing some of their sub-routines. Y. Zheng, S. Lu, and R. -B. Wu [9] designed an algorithm to train a perceptron using Grover's Algorithm [5]. M. Schuld, I. Sinayskiy, and F. Petruccione [10] developed a procedure that uses a quantum phase estimation algorithm in the classical neural network training process. P. Date, C. Schuman, R. Patton, and T. Poto [11] presents a Classical-Quantum Hybrid Approach for training unsupervised probabilistic machine learning models.

Machine learning (ML) studies techniques to give machines the ability to learn from past experiences. Its core tasks include classification or regression in supervised learning and density estimation or clustering in unsupervised learning, for example. Usually, in the ML context, the training of a machine is performed using a learning algorithm that uses as input a training data set [12] to extract information, adjust its parameters, and solve a given problem.

One of the ML techniques widely employed for many problems is the Artificial Neural Network (ANN), wherein the classical (non-quantum) world is a Classical ANN. Classical ANNs are one of the best-known classifiers and predictors. Classical ANNs have proven to be very competitive in solving real-world problems compared to other conventional data analysis methodologies [13]. Its optimization is observed from various perspectives, but in general, classical ANN training is performed mainly using a gradient descent algorithm. However, optimization methods based on the

• The authors are with the Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Recife, Pernambuco 52171-900, Brasil. E-mail: {lucianoserafimdesouza, hcarvalho.jon}@gmail.com, tiago.ESPNOLA@ufrpe.br.

Manuscript received 20 Apr. 2020; revised 30 Dec. 2020; accepted 3 Jan. 2021. Date of publication 13 Jan. 2021; date of current version 14 Jan. 2022. (Corresponding author: Luciano S. de Souza.)

Recommended for acceptance by J. C. Hoe.

Digital Object Identifier no. 10.1109/TC.2021.3051559

descending gradient algorithm have limitations. Aspects such as weight initialization, network architecture, activation functions, meta parameters, and learning environment can influence the optimization process [14], [15]. In this way, the training algorithm may not obtain a set of weights that train the neural network, staying stagnated at local minimums.

Therefore, a neural network's training algorithm based on the descending gradient method can be seen as a search problem and seeks to minimize an error function. This ANN training is a search problem for an appropriate weight configuration that allows learning the network [16]. On the other hand, in the quantum computing branch, some search algorithms are more efficient than their classical analogue. For example, Grover's quantum search algorithm [5] and quantum walks [17], [18], [19], [20], [21]. In this perspective, quantum search algorithms have already been used for the training process of classical ANN [9], [10].

Based on the incipient work developed by L. S. de Souza, J. H. A. de Carvalho, and T. A. E. Ferreira [22] with extreme learning machines, this article generalizes and extends the quantum walk proposal to train a classical ANN. Here, a full graph lackadaisical quantum walk algorithm [17], [19], [21] is applied as a search method to find all the synaptic weights that optimize the learning procedure of a classical ANN.

This paper is organized as follows. Section 2 introduces some concepts of the one-dimensional and the generalization for n -dimensional quantum walks, and it also presents the quantum walk on a complete graph. Section 3, it shows the computational procedure used in this paper. Section 4, it shows the experiments performed. Section 5 discusses the results obtained. Finally, Section 6 is the conclusion of the work.

2 QUANTUM WALK

The simplest model of the classical random walk can be described by a particle's classical movement in a straight line [23]. Let the s the particle probability of going to the right. Let $(1 - s)$ the probability of going to left. Therefore, the direction of the particle is conditioned by tossing a coin. This process is probabilistic, so it is impossible to know with certainty where the particle will be at any given time. However, it is possible to calculate the probability p that particle is at a point n at time t , as shown in Equation (1) for the case $s = 1/2$

$$p(t, n) \simeq \frac{2}{\sqrt{2\pi t}} e^{-\frac{n^2}{2t}}. \quad (1)$$

The quantum walks generalize the concept of a classical random walk, i.e., the quantum mechanical counterpart of classical random walks [24]. It is assuming a walker represented by a normalized vector in Hilbert space. The quantum walk evolves in the Hilbert space $\mathcal{H}_M \otimes \mathcal{H}_P$, where \mathcal{H}_M is the coin space that controls the walker's movement, and \mathcal{H}_P defines the position of the walker [23].

Suppose the quantum walk takes place in a one-dimensional space. In this case, the coin space needs two degrees of freedom. Therefore, the coin space \mathcal{H}_M is generated by the computational base $\{|0\rangle, |1\rangle\}$. A qubit can represent this

quantum information. The possible states that one qubit can assume are represented by the state vectors $|0\rangle$ and $|1\rangle$ described in Equations (2) and (3), respectively

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad (2)$$

and

$$|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (3)$$

Walker space \mathcal{H}_P is generated by base $\{|n\rangle : n \in \mathbb{Z}\}$ which represents all integers of one-dimensional space. Consider an operator S that, when applied to the system formed here by a coin and a walker, will shift the position of the individual to $|n+1\rangle$ or $|n-1\rangle$ depending on the state of the coin according to Equation (4)

$$\begin{aligned} S|0\rangle|n\rangle &= |0\rangle|n+1\rangle \\ S|1\rangle|n\rangle &= |1\rangle|n-1\rangle. \end{aligned} \quad (4)$$

A unitary transformation describes the evolution of a closed quantum system. This evolution depends on the application of an operator U shown in Equation (5) to the system over time [25]

$$U = S(H \otimes I), \quad (5)$$

where H is the Hadamard operator and I is the identity.

For an operator U of a quantum system to be unitary, it must satisfy the condition described in Equation (6), where U^\dagger is the adjunct of U . This condition is necessary for the norm of the vectors to be maintained, for this we must calculate $\|u\| = \sqrt{\langle u|u \rangle}$, where $\langle u|u \rangle$ is the inner product

$$UU^\dagger = U^\dagger U = I. \quad (6)$$

The evolution of the quantum walk system is performed by the Equation (7)

$$|\Psi(t)\rangle = U^t|\Psi(0)\rangle, \quad (7)$$

where $|\Psi(0)\rangle$ is the initial state of the quantum system.

Consider the initial state of the quantum system given by the Equations (8) and (9). Asymmetrical and symmetrical state, respectively. We can obtain the asymmetrical and symmetrical probability distributions after one hundred applications of U operator, shown in Figs. 1 and 2. Unlike the classical case, where the distribution is an origin-centered Gaussian, for both case asymmetrical and symmetrical, the quantum walk has a large spread with an interval of $-t/\sqrt{2}$ to $t/\sqrt{2}$

$$|\Psi(0)\rangle = |0\rangle|n=0\rangle \quad (8)$$

$$|\Psi(0)\rangle = \frac{|0\rangle - i|1\rangle}{\sqrt{2}}|n=0\rangle. \quad (9)$$

For example, if the quantum walk starts in the initial state described in Equation (8). Apply the Hadamard operator H as the coin of the quantum walk system and then applies the shift operator S . We have the initial evolution state presented in Equation (10) at the end of the first step.

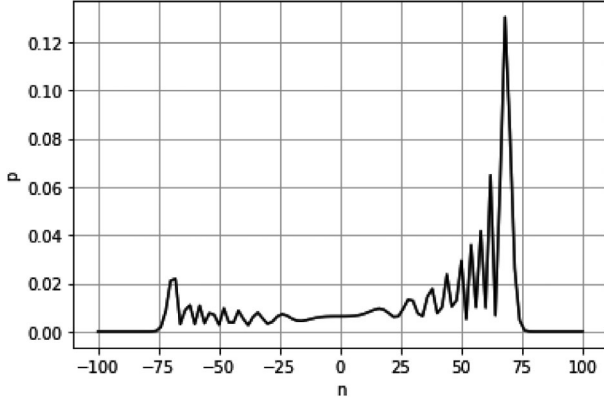


Fig. 1. Probability distribution of the one-dimensional quantum walk after 100 steps. The initial state is described by the Equation (8). The points with zero value were ignored.

The Hadamard operator application in the computational base vectors generates a superposition state, i.e., one of the most well-known quantum effects that qubits are at the same time in distinct states

$$|\Psi(1)\rangle = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|-1\rangle). \quad (10)$$

With the successive application of the evolution operator U described in Equation (5), at the end of the third stage, we can observe in Equation (11) that the state $|\Psi(3)\rangle$ is asymmetrical about the origin. This asymmetry will keep for all system evolution, as can be observed in Fig. 1

$$\begin{aligned} |\Psi(2)\rangle &= \frac{1}{2}(|0\rangle|2\rangle + (|1\rangle + |0\rangle)|0\rangle - |1\rangle|-2\rangle) \\ |\Psi(3)\rangle &= \frac{1}{2\sqrt{2}}(|0\rangle|3\rangle + (2|0\rangle + |1\rangle)|1\rangle \\ &\quad - |0\rangle|-1\rangle + |1\rangle|-3\rangle). \end{aligned} \quad (11)$$

2.1 One-Dimensional Quantum Walk

This section will describe the one-dimensional quantum walk model analytically and recursively. The generic state for a one-dimensional quantum walk model is described in Equation (12) [23].

$$|\Psi(t)\rangle = \sum_{n=-\infty}^{\infty} (\alpha_n(t)|0\rangle + \beta_n(t)|1\rangle)|n\rangle, \quad (12)$$

where the coefficients $\alpha_n(t)$ and $\beta_n(t)$ satisfy the condition described in Equation (13).

$$\sum_{n=-\infty}^{\infty} |\alpha_n(t)|^2 + |\beta_n(t)|^2 = 1. \quad (13)$$

Applying the operator $H \otimes I$ to state $|\Psi(t)\rangle$, we find recursive formulas involving the coefficients α and β in Equations (14) and (15).

$$\alpha_n(t+1) = \frac{\alpha_{n-1}(t) + \beta_{n-1}(t)}{\sqrt{2}} \quad (14)$$

$$\beta_n(t+1) = \frac{\alpha_{n+1}(t) - \beta_{n+1}(t)}{\sqrt{2}}. \quad (15)$$

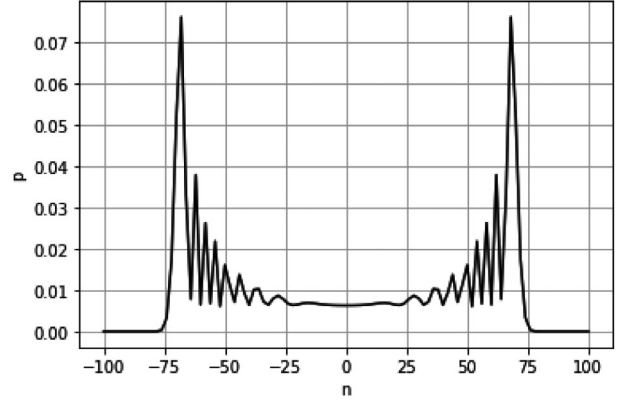


Fig. 2. Probability distribution of the one-dimensional quantum walk after 100 steps. The initial state is described by the Equation (9). The points with zero value were ignored.

As already shown in Figs. 1 and 2, the quantum walk probability distributions are dependent on the initial state. Therefore, the initial state of the system can generate a walk with symmetrical or asymmetrical probability distribution about the origin. The probability distribution can be calculated using the Equation (16)

$$p(t, n) = |\alpha_n(t)|^2 + |\beta_n(t)|^2. \quad (16)$$

2.2 n-Dimensional Quantum Walk

It is possible to generalize the concept presented in Section 2.1 to any number of dimensions. Consider the quantum walk in an infinite n -dimensional grid with the associated Hilbert space $\mathcal{H}_M \otimes \mathcal{H}_P$, whose \mathcal{H}_P 's computational base is $\{|x, y, \dots, n\rangle : x, y, \dots, n \in \mathbb{Z}\}$, and the Coin space \mathcal{H}_M 's computational basis is $\{|i_x, i_y, \dots, i_n\rangle : i_x, i_y, \dots, i_n \in \{0, 1\}^n\}$. The generic state for this quantum walk model at time t is presented in Equation (17).

$$\begin{aligned} |\Psi(t)\rangle &= \sum_{i_x, i_y, \dots, i_n=0}^1 \sum_{x, y, \dots, n=-\infty}^{\infty} \\ &\quad \psi_{i_x, i_y, \dots, i_n; x, y, \dots, n}(t) \\ &\quad |i_x, i_y, \dots, i_n\rangle |x, y, \dots, n\rangle, \end{aligned} \quad (17)$$

where $\psi_{i_x, i_y, \dots, i_n; x, y, \dots, n}(t)$ are complex functions that satisfy the condition shown in Equation (18) for all time t

$$\sum_{i_x, i_y, \dots, i_n=0}^1 \sum_{x, y, \dots, n=-\infty}^{\infty} |\psi_{i_x, i_y, \dots, i_n; x, y, \dots, n}(t)|^2 = 1. \quad (18)$$

Moreover, it is possible to calculate the distribution of probabilities using Equation (19).

$$p_{x, y, \dots, n}(t) = \sum_{i_x, i_y, \dots, i_n=0}^1 |\psi_{i_x, i_y, \dots, i_n; x, y, \dots, n}(t)|^2. \quad (19)$$

Applying the standard evolution operator $U = S(C \otimes I)$ (C is the coin operator) to the generic state described in Equation (17) and making the expansions, we obtain Equation (20) which is the walker evolution equation [23]

$$\begin{aligned}
& \psi_{i_x, i_y, \dots, i_n; j_x, j_y, \dots, j_n}(t+1) \\
&= \sum_{j_x, j_y, \dots, j_n=0}^1 C_{i_x, i_y, \dots, i_n; j_x, j_y, \dots, j_n} \\
& \psi_{j_x, j_y, \dots, j_n; x+(-1)^{i_x}, y+(-1)^{i_y}, \dots, n+(-1)^{i_n}}(t).
\end{aligned} \tag{20}$$

2.3 Lackadaisical Quantum Walk on Complete Graph

The quantum walks presented in the previous sections are the basis for other techniques that can be used in other search spaces. One such variation is the quantum walk in a complete graph developed by Wong [17], represented in Fig. 3.

There are two types of vertex, a and b , marked as a solution and non-solution, respectively. Each one vertex has l self-loops. This approach considers the walker's movement into the complete graph to create the states of the new computational basis.

If the walker is on an a vertex, there are two movement options. It can move to a vertex that is a solution ($a \rightarrow a$) or to a vertex that is not a solution ($a \rightarrow b$). Defining the quantum states $|a\rangle \otimes |a \rightarrow a'\rangle$ and $|a\rangle \otimes |a \rightarrow b\rangle$. Similarly, if the walker is on any vertex b . It can move to a vertex that is a solution ($b \rightarrow a$) or to a vertex that is not a solution ($b \rightarrow b'$). Defining the quantum states $|b\rangle \otimes |b \rightarrow a\rangle$ and $|b\rangle \otimes |b \rightarrow b'\rangle$.

Note that the states $|a\rangle$ or $|b\rangle$ represent the graph vertices and $|a \rightarrow b\rangle$, for example, represents the edges where $|a\rangle$ is the walker's current state and $|b\rangle$ is the state to which the walker will move to it. Equation (21) shows the states $|AA\rangle, |AB\rangle, |BA\rangle$, and $|BB\rangle$ of the new quantum states for the situation where there is only one solution (only one vertex labeled as a). N is the total number of vertices. Since there is only one solution, $k = 1$, and the number of self-loops is greater than zero, $l > 0$.

$$\begin{aligned}
|AA\rangle &= \frac{1}{\sqrt{l}} |a\rangle \otimes |a \rightarrow a'\rangle \\
|AB\rangle &= \frac{1}{\sqrt{N-1}} \sum_b |a\rangle \otimes |a \rightarrow b\rangle \\
|BA\rangle &= \frac{1}{\sqrt{N-1}} \sum_b |b\rangle \otimes |b \rightarrow a\rangle \\
|BB\rangle &= \frac{1}{\sqrt{(N-1)(N+l-2)}} \sum_b \sum_{b'} |b\rangle \otimes |b \rightarrow b'\rangle.
\end{aligned} \tag{21}$$

Equation (22) defines the initial state $|\Psi_0\rangle$. This state is described as the uniform superposition $N^{-1} \sum_{x,y} |x\rangle |x \rightarrow y\rangle$ expressed in terms of the states AA, AB, BA and BB

$$\begin{aligned}
|\Psi_0\rangle &= \frac{1}{\sqrt{N(N+l-1)}} (\sqrt{l}|AA\rangle + \sqrt{N-1}|AB\rangle \\
&+ \sqrt{N-1}|BA\rangle + \sqrt{(N-1)(N+l-2)}|BB\rangle).
\end{aligned} \tag{22}$$

The lackadaisical quantum walk is accomplished by successive applications of a unitary operator U , defined in Equation (23) which inverts the sign of solution states $|a\rangle |a \rightarrow x\rangle$ using an oracle and swaps vertices on each edge, $|x\rangle |x \rightarrow y\rangle \rightarrow |y\rangle |y \rightarrow x\rangle$ as described in Wong [17]

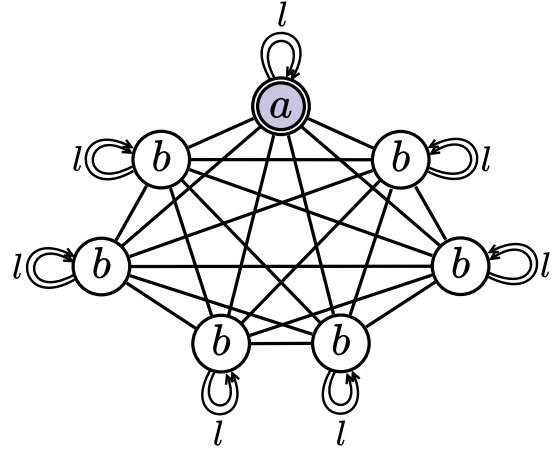


Fig. 3. Complete graph with $N = 7$ vertices. The single vertex marked as a is indicated by the double circle shaded. Adapted from Wong's work [17]. Grover search with lackadaisical quantum walks.

$$U = \begin{pmatrix} \cos \theta & -\sin \theta & 0 & 0 \\ 0 & 0 & -\cos \phi & \sin \phi \\ -\sin \theta & -\cos \theta & 0 & 0 \\ 0 & 0 & \sin \phi & \cos \phi \end{pmatrix}, \tag{23}$$

where θ is defined such that

$$\cos \theta = \frac{N-l-1}{N+l-1},$$

and

$$\sin \theta = \frac{2\sqrt{l(N-1)}}{N+l-1},$$

and ϕ is defined such that

$$\cos \phi = \frac{N-l-3}{N+l-1},$$

and

$$\sin \phi = \frac{2\sqrt{N+l-2}}{N+l-1}.$$

Thus, the evolution of the system occurs in a four-dimensional subspace, and each state of the new representation is formed by overlapping vertices and edges [17].

2.3.1 Quantum Walk With Self-Loops for k Solutions

The previous section introduced the quantum walk in a complete graph and describes the approach to the case of a single solution ($k = 1$) and the number of self-loops $l > 0$. Now, consider a number of solutions $k > 1$ and only one self-loop ($l = 1$) per vertex.

The solution set has the number k of vertices, and the non-solution set has $N - k$ vertices, where N is the total number of vertices. We will use the previous idea that considered the walker's movement to define the states of the new multiple solutions computational basis.

If the walker is in a vertex a moving to another vertex a' , there will exist k edges of the type $|a \rightarrow a'\rangle$. If the walker is in a vertex a moving to a vertex b , there will exist $N - k$

edges of the type $|a \rightarrow b\rangle$. Similarly, if the walker is in a vertex b moving to a vertex a , there will exist k edges of the type $|b \rightarrow a\rangle$. And if walker is in a vertex b moving to another vertex b' , there will exist $N - k$ edges of the type $|b \rightarrow b'\rangle$.

The number of vertices marked as solution must be of the order $o(N)$ because if $k = O(N)$, then $k = cN$ (in the limit of $N \rightarrow \infty$ and c is a finite constant). In this last case, the search for one solution could be performed classically in an efficient way, i.e., in a constant number of guesses [17]. Thus, the new quantum states are redefined in Equation (24).

$$\begin{aligned} |AA\rangle &= \frac{1}{k} \sum_a \sum_{a'} |a\rangle \otimes |a \rightarrow a'\rangle \\ |AB\rangle &= \frac{1}{\sqrt{k(N-k)}} \sum_a \sum_b |a\rangle \otimes |a \rightarrow b\rangle \\ |BA\rangle &= \frac{1}{\sqrt{k(N-k)}} \sum_b \sum_a |b\rangle \otimes |b \rightarrow a\rangle \\ |BB\rangle &= \frac{1}{N-k} \sum_b \sum_{b'} |b\rangle \otimes |b \rightarrow b'\rangle. \end{aligned} \quad (24)$$

The initial state of the system is rewritten, as presented in Equation (25).

$$\begin{aligned} |\Psi_0\rangle &= \frac{1}{N} (k|AA\rangle + \sqrt{k(N-k)}|AB\rangle \\ &\quad + \sqrt{k(N-k)}|BA\rangle + (N-k)|BB\rangle). \end{aligned} \quad (25)$$

A modification to the evolution operator U (Equation (23)) is made so that the number of solutions is included in the definitions of θ and ϕ [17]. Therefore, θ is redefined according to Equations (26) and (27).

$$\cos \theta = \frac{N - 2k - l + 1}{N + l - 1} \quad (26)$$

$$\sin \theta = \frac{2\sqrt{(N-k)(k+l-1)}}{N + l - 1}, \quad (27)$$

and ϕ is redefined according to Equations (28) and (29).

$$\cos \phi = \frac{N - 2k + l - 1}{N + l - 1} \quad (28)$$

$$\sin \phi = \frac{2\sqrt{k(N-k+l-1)}}{N + l - 1}. \quad (29)$$

The maximum success probability value is reached after the number of steps t defined in Equation (30) [17]. Success is defined as the measurement of some state $|AA\rangle$ or the state $|AB\rangle$. Both states represent the set of vertices marked as a solution.

$$t = \frac{\pi}{\sqrt{2(2k+l-1)}} \sqrt{N}. \quad (30)$$

2.3.2 A Toy Example

Consider the particular case of the lackadaisical quantum walk in a complete graph with $N = 8$ vertices, $k = 2$

solutions, and $l = 1$ self-loops at each vertex. For illustrative purposes, these vertices have also been marked with sub-indices, so a_1 and a_2 are the solutions whereas vertices b_1, \dots, b_6 are not solutions. In practice, however, sub-indices or whatever kind of information that can distinguish the solutions from one another, or the non-solutions from one another, are not available. The vertices are marked only with a or b , exclusively.

The quantum state $|AB\rangle$, for example, is formed by all vertices that are solution combined with their respective edges for non-solution vertices. Therefore, the state $|AB\rangle$ for the given example is defined according to Equation (31).

$$\begin{aligned} |AB\rangle &= \frac{1}{\sqrt{k(N-k)}} \sum_a \sum_b |a\rangle \otimes |a \rightarrow b\rangle \\ &= \frac{1}{\sqrt{12}} \left[\left(|a_1\rangle |a_1 \rightarrow b_1\rangle + |a_1\rangle |a_1 \rightarrow b_2\rangle \right. \right. \\ &\quad + |a_1\rangle |a_1 \rightarrow b_3\rangle + |a_1\rangle |a_1 \rightarrow b_4\rangle \\ &\quad + |a_1\rangle |a_1 \rightarrow b_5\rangle + |a_1\rangle |a_1 \rightarrow b_6\rangle \Big) \\ &\quad + \left(|a_2\rangle |a_2 \rightarrow b_1\rangle + |a_2\rangle |a_2 \rightarrow b_2\rangle \right. \\ &\quad + |a_2\rangle |a_2 \rightarrow b_3\rangle + |a_2\rangle |a_2 \rightarrow b_4\rangle \\ &\quad \left. \left. + |a_2\rangle |a_2 \rightarrow b_5\rangle + |a_2\rangle |a_2 \rightarrow b_6\rangle \right) \right]. \end{aligned} \quad (31)$$

The analysis for the quantum states $|AA\rangle$, $|BA\rangle$ and $|BB\rangle$ is analogous. With the new quantum states prepared, the initial state of the system $|\Psi_0\rangle$ can be defined according to Equation (32). This initial state is also normalized.

$$|\Psi_0\rangle = \frac{1}{8} \left(2|AA\rangle + \sqrt{12}|AB\rangle + \sqrt{12}|BA\rangle + 6|BB\rangle \right). \quad (32)$$

Making only three ($t = 3$) successive applications of the evolution operator U , represented by the matrix of Equation (33), the probability of success (the measurement of a solution state) tends to 1. A measurement made in $|\Psi_3\rangle$ makes the system to collapse to the state $|AA\rangle$. All energy of the system is concentrated at this state $|AA\rangle$.

$$U = \frac{1}{2} \begin{pmatrix} 1 & -\sqrt{3} & 0 & 0 \\ 0 & 0 & -1 & \sqrt{3} \\ -\sqrt{3} & -1 & 0 & 0 \\ 0 & 0 & \sqrt{3} & 1 \end{pmatrix}. \quad (33)$$

3 PROPOSAL PROCEDURE

Quantum walks are algorithms that can be applied to search problems [19], [20], [26], where the proof of its correctness can be found in [27]. Thus, it is possible to idealize this algorithm's application to find the set of synaptic weights that train a classical artificial neural network. Based on the concept of quantum information processing called agent-environment presented in Section 1, the objective is to replace the classical algorithm with a quantum search algorithm in a neural network training process.

In a classical environment, we replace the backpropagation algorithm for a quantum walk to search synaptic

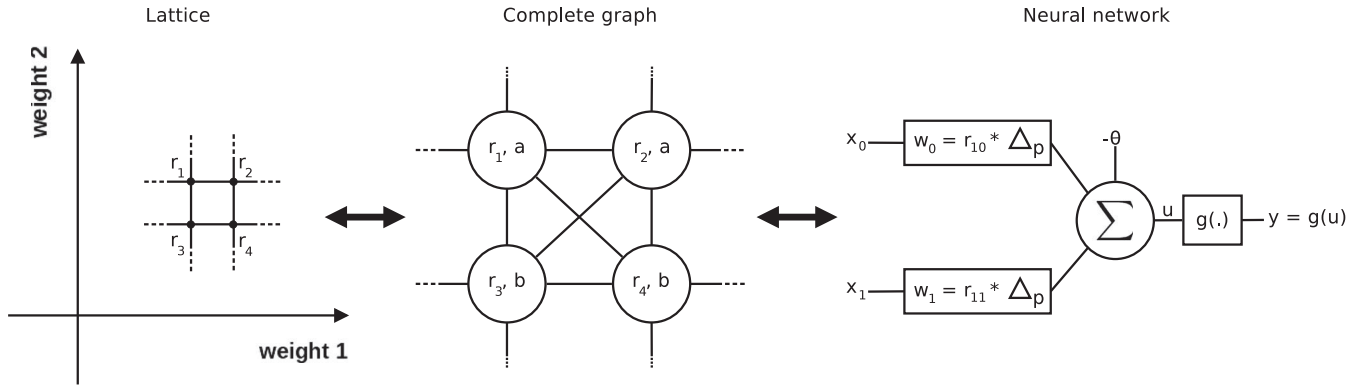


Fig. 4. On the left side, a discrete schematic representation of the set of weight values of an artificial neural network. This representation generates a grid, represented here by a lattice. Each intersection point i of the lattice has the information about the set of weight values r_i . With an oracle operator, the state solution label (a for a solution and b for no solution) is determined. On the central figure, the complete graph representation of the grid. On the right side, a neural network with one neuron demonstrates the relationship between the labels on the grid, the complete graph, and the definition of synaptic weights w_0 and w_1 , where $g(\cdot)$ is the activation function, and $u = x_0 \cdot w_0 + x_1 \cdot w_1 - \theta$.

weights. The evolution of the quantum walk occurs by applying a unitary operator U over an initial state. Subsequently, a measurement of the walker state is performed. If the measurement is performed at each step, then the quantum walk falls in the classical case. In this situation, the correlations between the different positions of the walker are lost.

For the correlations between the walker's positions to be maintained, the measurement should not occur at every single step. Thus, the measurement process only will occur after a predetermined time t . Once the correlations between positions hold, constructive and destructive interference occurs [23]. The interference caused by these relationships between positions generates the probability amplitudes for each position. Depending on the initial state of the system, the probability distribution may be asymmetrical or symmetrical according to Equations (8) and (9), respectively, as viewed in Section 2. As observed in Figs. 1 and 2, according to the probability distribution, the chances of finding the walker on the extremities are higher than finding it in another position. In this way, it is possible to obtain these extreme positions of the walker with a high probability of measurement.

Therefore, it is possible to determine with a high probability the walker's extreme position after a given number of steps. If this position does not contain a solution, the quantum walk search will not succeed. It is necessary to know the position where there is a solution and guarantee an amplitude amplification for this position. It is also necessary to guarantee that the solution state at the time of measurement has a high probability of measurement. The complete-graph quantum walk proves to be adequate for these proposals. The complete-graph quantum walk employed here considers only the information about the *solution* or *no solution* label (a or b) contained in all vertices to execute the quantum walk evolution. However, to define these labels, the information about the weights used to train the classical ANN is necessary. Thus, besides the label for *solution* ($|a\rangle$) and *no solution* ($|b\rangle$), each vertex will also have the information about the associated weights used to train the classical ANN.

Consider a w -dimensional lattice. Let a classical ANN with the number of weights equal to the lattice's dimension, w weights. In the discrete representation, each intersection of lattice lines can represent an ANN w weights configuration.

Fig. 4 represents this idea for the 2-dimensional case. In this way, each point i in this w -dimensional lattice will be a vertex in a complete graph, where it was labeled as a if its value configuration of weights r_i is a solution for the ANN. Otherwise, it was labeled as b . The labels a and b are created by applying the oracle to the grid. Change the grid representation to the complete graph creates a new representation, where all i grid point is a vertex with the information r_i (the weight set) and the label a or b . Wong's quantum walk [17] is applied in the complete graph, where now each vertex has the label a or b , and the associated weight set (r_i) used to train the classical ANN. Then, the search procedure is done, where a vertex labeled by $|a\rangle$ is searched.

Thus, to recover the synaptic weights that trained the ANN after the quantum walk evolution, it is necessary to obtain the specific weight values configuration r_i of the state measured. For this reason, a modification is proposed to the original procedure, including the weight vector $|\vec{r}_i\rangle$ in the base states definition. All quantum walk search procedure is the same, where the state with label a is sought. However, with this new associated information about the weight vector $|\vec{r}_i\rangle$, it is possible to determine the walker's position on the grid (or lattice) in the final measurement procedure in the final state found by the quantum walk algorithm. In this way, the information about the set of weight values used to train the classical ANN can be recovered. The new definition of states is presented in Equation (34). With this modification, it is proposed the search procedure presented in Algorithm 1.

$$\begin{aligned}
 |AA\rangle &= \frac{1}{k} \sum_a \sum_{a'} |\vec{r}_a\rangle |a\rangle \otimes |a \rightarrow a'\rangle \\
 |AB\rangle &= \frac{1}{\sqrt{k(N-k)}} \sum_a \sum_b |\vec{r}_a\rangle |a\rangle \otimes |a \rightarrow b\rangle \\
 |BA\rangle &= \frac{1}{\sqrt{k(N-k)}} \sum_b \sum_a |\vec{r}_b\rangle |b\rangle \otimes |b \rightarrow a\rangle \\
 |BB\rangle &= \frac{1}{N-k} \sum_b \sum_{b'} |\vec{r}_b\rangle |b\rangle \otimes |b \rightarrow b'\rangle.
 \end{aligned} \tag{34}$$

Initially, it is necessary to define some initial parameters of the Algorithm 1. A w -dimensional grid (or a lattice) will

define the possible weight configurations for the walker, where each dimension represents an ANN weight set. Therefore, the user defines the number of grid points, N , and the distance between adjacent points, Δp . In this way, N will be the number of possible positions state of the system for the walker, represented by the complete graph, and Δp will define the granularity of the weights representation. If Δp is small, the weight representation will have high resolution, but a significant N value is necessary to search for a solution in practice. If Δp is large, the weight search will have low resolution, but a small N value is sufficient to cover a given search space. The algorithm employs an oracle, which defines the vertices that are solution and non-solution. The oracle employed here is described in Section 4.5. In this way, it is considered that the oracle used in this work is independent of the algorithm proposed. An w -dimensional sparse matrix with N elements represents classically the oracle, where the element marked with number one denotes a solution and with zero a non-solution. The user also defines the number of self-loops per vertex, l . However, the number of self-loops per vertex here always was $l = 1$ for all experiments.

Algorithm 1. Training Algorithm

```

1 begin
2   Set the parameters:  $\Delta p$ ,  $N$  and  $l$ 
3   Quantum count execution
4   Preparation of the initial state
5   for  $j \leftarrow 1$  to  $\frac{\pi}{\sqrt{2(2k+l-1)}}\sqrt{N}$  do
6      $|\Psi_j\rangle \leftarrow U|\Psi_{j-1}\rangle$ 
7   end
8   Make the measurement
9   Initialize the weights of the Artificial Neural Network
10 end

```

After defining the initial parameters, the proposed algorithm performs a quantum count to estimate the number of solutions k in the search space, as indicated in line 3 of the Algorithm 1. It is necessary to know the quantity k of solutions to determine the initial state, the shift operator U , and the number t of iterations. It is possible to define the number of solutions by combining the phase estimation technique based on the Fourier Quantum Transform with the Grover iteration [25]. Alternatively, it is also possible to apply the amplitude estimation to the problem of approximate counting [28], or to use techniques inspired by Shor's celebrated quantum factorization algorithm and combines them with Grover's algorithm [29]. Note that the number of solutions is known at this moment, but their search space positions are not known.

With the number k of solutions determined, the states will be constructed according to Equation (34). The quantum walk occurs in a complete graph, as shown in Fig. 3, where the quantum walk will be performed in a four-dimensional space according to in Section 2.3, reducing the search space. Thus, the search space is represented by the superposition of vertices and edges already presented in Equation (34).

Following the Algorithm 1 in line 4, the initial state preparation is performed considering the space size, the number

of solutions, and the relations between vertices and edges, according to Equation (25).

After preparing the system, the quantum walk is performed. Line 5 of the Algorithm 1 defines the total number of steps t , according to Equation (30). At each step j , the evolution operator U is applied to the quantum system $|\Psi_{j-1}\rangle$, where $j = 1, 2, \dots, t$, as shown in line 6.

Once the evolution is completed, a measurement in the basis $|x\rangle|x \rightarrow y\rangle$ is performed, as presented in line 8. Thus, the states $|a\rangle|a \rightarrow a'\rangle$ or the states $|a\rangle|a \rightarrow b\rangle$, which both have the solutions, is obtained with high probability. It is worth noting that the vertex information where the walker stays defines the found solution at the time of the measurement. The direction where the walker points for the next quantum movement (defined by its edge) is only relevant to define the quantum walk evolution, not to define the solution state at the measurement time.

After recovering the states $|a\rangle|a \rightarrow a'\rangle$ or $|a\rangle|a \rightarrow b\rangle$ with high probability after the measurement, in the line 9, the algorithm initializes the classical neural network weights as shown on the right side of Fig. 4.

4 EXPERIMENT SETUP

A simple classification problem was performed to assess the concept of the proposed algorithm. An artificial neural network of MLP type (Multilayer Perceptron Type) was employed to solve the "EXCLUSIVE-OR" classification problem. The neural network was initialized with the weights generated by the procedure proposed in this work.

4.1 Exclusive-OR Function

The EXCLUSIVE OR (XOR) problem, a simple but non-linearly separable problem, was used as the function to be learned by the neural network. The XOR function, also known as exclusive disjunction, is an operation on two binary values, x_0 and x_1 , where if only one of these binary values is equal to 1, then the function returns 1, otherwise returns 0 [30].

Consider the problem of classifying points in the unitary hypercube. The EXCLUSIVE OR can be understood as a particular case of this problem. In this case, it is sufficient to consider only the four vertices of the unit square corresponding to the points $\{(0,0), (0,1), (1,0), (1,1)\}$. Each set of patterns determine outputs that are called classes. The inputs $\{(0,0), (1,1)\}$ generate outputs $x_0 \oplus x_1 = 0$, where it will be called class 0. The inputs $\{(0,1), (1,0)\}$ generate outputs $x_0 \oplus x_1 = 1$, class 1 [14]. Therefore, it is a pattern classification problem that consists of associating an input pattern (x_0, x_1) with one of the previously defined classes $\{0, 1\}$ [31].

4.2 Neural Network Architecture

The EXCLUSIVE OR problem can be solved by a Multi-Layer Perceptron (MLP) neural network with three neurons, where two of them are in the hidden layer, and the other neuron is in the output layer. The neural network used in this work follows this architecture.

The input layer has two values $\{x_0, x_1\}$ that are inputs data. The hidden layer has two neurons. The output layer has one neuron. Each neuron has a bias. Thus, the neural

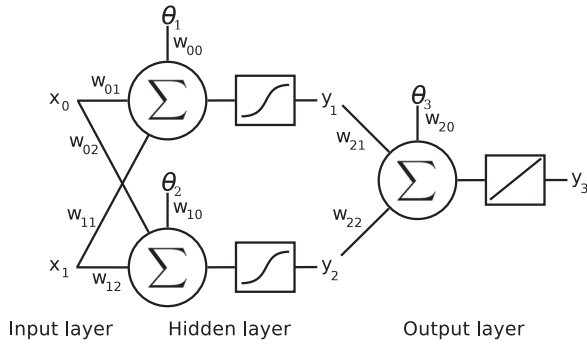


Fig. 5. Configuration of a multilayer neural network. The letters θ_1 , θ_2 and θ_3 mean the bias and the letters y_1 , y_2 , and y_3 mean the outputs of neurons.

network has nine synaptic weights. Six weights are in the hidden layer $\{\omega_{00}, \omega_{01}, \omega_{02}, \omega_{10}, \omega_{11}, \omega_{12}\}$ and three are in the output layer $\{\omega_{20}, \omega_{21}, \omega_{22}\}$ according to Fig. 5. The sigmoid logistic, described by Equation (35), is the activation function for all neurons of the hidden layer and the neuron of the output layer is linear

$$f(x) = \frac{L}{1 + e^{-\lambda(x-\zeta_0)}}, \quad (35)$$

where $\lambda = 1$ is the declivity of the curve, $L = 1$ is the maximum value of the curve, and $\zeta_0 = 0$ is the value of x at the midpoint of the curve.

4.3 Hardware and Software Setup

The simulations were performed using the following hardware configurations. For simulate neural network training with classical backpropagation algorithm, it was used,

- Operational System: Debian GNU/Linux 10 Buster;
- Memory: 4 GiB;
- Processor: Intel Core i3-5005U CPU @ 2.00 GHz x 4;
- OS type 64-bit;
- HD: 1 terabytes.

To simulate the procedure using a quantum walk in the complete graph, it was used

- Operational system: Debian GNU/Linux Jessie 8.11;
- Memory: 16 GiB;
- Processor: Xeon Intel 5th gen CPU @ 3.6 GHZ x 8;
- OS type 64-bit;
- HD: 4 terabytes.

The programming languages used to write the algorithms were Python 3.6 with the open-source machine learning framework PyTorch.¹

4.4 Search Space

Theoretically, the quantum walk is performed in an infinite space. However, in practice, because of memory and hardware limitations, we have determined sub-spaces or windows. Once the window size is defined, the procedure performs the search within it.

Geometrically, the windows employed here will always be hyper-cubes in the search space. For example, in the 2-dimensional case, a window is a square with $N = z^2$ points,

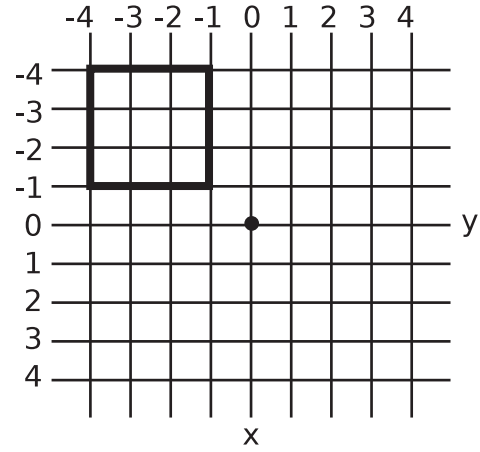


Fig. 6. Illustration of a search in an infinite two-dimensional grid using a window with a finite set of points.

where z is the number of points in a dimension. For the d -dimensional case, a window is a hyper-cube with $N = z^d$ points, where there are z points in each dimension.

If there is no solution in the sub-region defined by the windows, the windows will be shifted. By applying offsets, the windows are moved by performing the search in the infinite search space regions. To illustrate, consider an infinite two-dimensional search space as illustrated in Fig. 6. These window shifts in infinite space are carried out until at least one solution is found in the current window. There are many forms to define the shift of the windows. However, the simplest way is to sum an offset of size z for each dimension with respect to the current window, which was the approach used in this work.

4.5 Representation of the Oracle

An oracle is a structure capable of generating answers to binary questions. The circuit shown in Fig. 7 represents the implementation of the Deutsch algorithm. Conditional port U_f implements the NOT-controlled port with control bit $f(x)$ and acts as $U_f : |x, y\rangle \rightarrow |x, y \oplus f(x)\rangle$.

Port U_f is a black box with no explicit implementation, often called an oracle [32]. In this way, a quantum oracle is a “black box” operator that, when applied to a system state, return if this specific state is a solution or is not a solution. An example of oracle implementation can be seen in the work of Yu Zheng, Sicong Lu, and Re-Bing Wu [9], which uses Grover’s oracle as a central part of the circuit that implements their proposal of quantum perceptron models [33].

In this way, it was necessary to create a representation of an oracle to perform the simulations. The solution employed uses a sparse matrix with the size of the search space. The positions filled with 1 indicate the points that are a solution and 0 otherwise. In practice, the oracle answers whether a given point in the walker space, after converted to synaptic weights, correctly classifies the input patterns or not. However, a formal conception of this quantum oracle for this

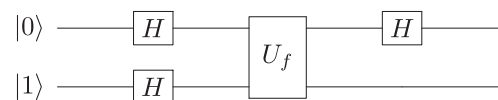


Fig. 7. The circuit that implements the Deutsch algorithm [4].

1. <https://pytorch.org/>

purpose is beyond the scope of this work. Therefore, we considered it existing.

4.6 Weight Generation

The quantum walk performs a search in quantum states that represent points in an integer space. At the end of the process, it is necessary to convert these points into synaptic weights values represented by real numbers. Thus, a real value was defined, Δp , where it will multiply each measured point at the end of the process. That is, the walker grid space is in Δp units. The walker starts at the window center, and the walker will move by the positive and negative integer indices of each dimension of the grid.

Therefore, when the proposed algorithm makes the measurement at the walker position in the grid $|r_i\rangle$ (where i is the label for each point in the w -dimensional search space), the integer components of $r_i = (n_1, n_2, \dots, n_w)$ for each dimension are converted to the real values $\Delta p * n_j$ ($j = 1, 2, \dots, w$), generating the synaptic weights for the neural network. In this way, the search is performed only among the factors of Δp . Also, the value defined for Δp establishes the search refinement level.

4.7 Measurement Process

Through unitary transformation, closed quantum systems evolve. In order to be able to access information that is in the $|\Psi\rangle$ state, an observation must be made. In practice, measurements are made in laboratories using physical devices, such as lasers, magnets, scales, and stopwatches, but in theory, this measurement process is described mathematically [23]. According to the postulate of quantum mechanics, the probability of a state occurs is

$$p(m) = \langle \Psi | M_m^\dagger M_m | \Psi \rangle, \quad (36)$$

where M is a measurement operator, \dagger is the symbol used to describe the conjugate transpose operation, and the index m represents the results that can occur in the experiment after the measurement. The state $|\Psi'\rangle$ described in Equation (37) is the state of the system just after the measurement [25]

$$|\Psi'\rangle = \frac{M_m |\Psi\rangle}{\sqrt{\langle \Psi | M_m^\dagger M_m | \Psi \rangle}}. \quad (37)$$

As an example we are going to measure the state $|\Psi\rangle$ whose states are described in Equation (34). Considering the system with four possible results we will define the operators $\{|AA\rangle\langle AA|, |AB\rangle\langle AB|, |BA\rangle\langle BA|, |BB\rangle\langle BB|\}$. So when applying Equations (36) and (37) we have

$$p(m) = |\lambda_m|^2$$

$$|\Psi'\rangle = \frac{\lambda_m}{|\lambda_m|} |m\rangle,$$

where $m = AA, AB, BA, BB$.

In this way, the measurement is a probabilistic process, where theoretically, the probability of measuring a given state $|m\rangle$ will be $|\lambda_m|^2$. Thus, after the quantum walk evolution, it is possible to observe from the final quantum state the quantity $|\lambda_m|^2$ for each state component $|m\rangle$.

Statistically, as we can see in Section 5, the chances of measuring the states that contain a solution were much more significant than measuring a state that did not contain a solution.

4.8 Comparison Against Backpropagation Algorithm

Since the search space contains solutions to the problem, the approach proposed in this paper can find a solution with high probability after the number of iterations defined in Equation (30). It implies that, in addition to ensuring a successful search, the number of iterations required to reach the artificial neural network training is known *a priori*.

The classical backpropagation algorithm seeks to minimize an error function based on the gradient descent methodology. For this reason, it can stagnate in local minimums. The practical implication is the possibility that the algorithm runs indefinitely and yet does not generate correct outputs. These limitations of a gradient descent methodology are overcome with the proposed quantum algorithm.

Furthermore, many experiments were done with both approaches to compare the proposed algorithm with the traditional backpropagation algorithm. The criterion to measure the performance was the number of iterations needed to find an ANN weight set that solves the XOR problem. The number of epochs for the backpropagation and the number of iterations for the proposed algorithm were observed.

Comparing the number of epochs of the backpropagation algorithm with the number of iterations of the procedure developed in this work was applied to measure how much better one approach is than the other. The efficiency of the proposed procedure is the order of $O(\sqrt{N/k})$. However, it was not possible to compare the efficiencies in terms of running time between our procedure and the backpropagation algorithm.

5 EXPERIMENTAL RESULTS AND DISCUSSIONS

A set of experiments with classical backpropagation was done to define a comparative baseline. Employing the same ANN architecture $2 - 2 - 1$ (two inputs, two hidden neurons, and one output – according to Section 4.2), five different learning rates $\eta = \{0.5, 0.1, 0.01, 0.001, 0.0001\}$ were investigated, where 1200 simulations were computed for each one. The maximum number of epochs of 150000, the training stagnation, and the zero classification error were the stopping conditions employed. The network was considered in a stagnation training situation if its MSE error did not decrease by 1000 consecutive epochs. The classification error is zero when the ANN can classify all the four XOR inputs correctly.

For these conditions, none of the classical backpropagation experiments reached a stagnation situation. The experiments with learning rates of 0.5 and 0.1 never reached the maximum number of epochs, always reaching the perfect classification. The summarization of this experimental behavior can be viewed in Table 1, where *lr* indicates the learning rates, *Epochs limit* is the number of times that the experiment reached the maximum number of epochs stop condition, and *Successful* is the number of experiments that obtained zero classification error.

TABLE 1
Classical Backpropagation Simulations

lr	Epochs limit	Successful
0.5000	None	1200
0.1000	None	1200
0.0100	452	748
0.0010	467	733
0.0001	726	474

The column *lr* presents the investigated learning rate, *Epochs limit* shows the experiment number that reached the maximum number of epochs, and *Successful* the number of experiments with zero classification error.

Table 2 presents the descriptive statistics for the baseline experiments. For each learning rate, it is presented the minimum, mean, maximum value of the Epochs number, and its standard deviation (Std.).

Table 3 presents the results of the simulations performed using the proposed computational procedure described in Algorithm 1. Here, k represents the number of solutions, N the number of vertices, l the number of self-loops (for all experiments $l = 1$), and t the number of iterations. $\Delta p = 0.5$ was utilized for all experiments. Here, it was investigated three configurations of N ($2^9 = 512$; $4^9 = 262144$; $8^9 = 134217728$) in five experiments. Two experiments with $N = 512$, 2 points per dimension. Two experiments with $N = 262144$, 4 points per dimension, and one experiment with $N = 134217728$, 8 points per dimension. The search spaces are sub-regions (windows) of an infinite ω -dimensional grid, as presented in Section 4.4. However, the initialization of these sub-regions can occur in different positions of the infinite grid. If there are no solutions in the current sub-region, the sub-region is moved to another region of the infinite grid. All simulations had their sub-region initiated randomly around the origin. The shift of the sub-region occurs until a region with at least one solution is found. With a few interactions, the sub-space converged to a region with solutions in all cases studied here.

Coincidentally, although the experiments 1 and 2 used distinct search windows, the proposed procedure converged to sub-spaces with 12 solutions (see Table 3). As $N = 512$ and $k = 12$ for those experiments, the iteration number t is also equal, being $t = 11$, which was theoretically defined to 10.26 by Equation (30). For the simulations 3 and 4, the proposed procedure converged to sub-spaces with 17 and 20 solutions, respectively. Thus, the iteration number t is different for those simulations, being 196 and 180, theoretically defined to 195.83 and 179.83, respectively. For the simulation 5, the

TABLE 2
Descriptive Statistics

lr	Epoch Statistics			
	Minimum	Mean	Maximum	Std.
0.5000	1	33.60	319	35.68
0.1000	3	433.84	3279	463.78
0.0100	2	5277.48	132199	17927.67
0.0010	9	12949.18	148256	22451.79
0.0001	295	46987.00	149644	36780.22

Parameters of the backpropagation algorithm experiment for the number of epochs. *lr* means the learning rate. *Std.* means Standard Deviation.

TABLE 3
Weight Search Experiment by the Quantum Walk Procedure

Experiment	k	N	t	
			Theoretical	Simulated
1	12	512	10.26	11
2	12	512	10.26	11
3	17	262144	195.83	196
4	20	262144	179.83	180
5	80295	134217728	64.22	65

K means the number of solutions. N is the total number of vertices. The number of iterations is represented for letter t . For all experiments, the number of self-loops was $l = 1$.

procedure converged to a sub-space with more than eighty thousand solutions (80295), which implied in $t = 65$ iterations (theoretical number of iterations of 64.22).

The procedure developed using the quantum search algorithm, even in spaces with a high number of vertices, was able to amplify the amplitudes in a relatively low number of iterations compared with the results for the classical backpropagation procedure. In the best case, the maximum iteration number for the quantum walk algorithm was 11, while the mean number of epochs obtained by the backpropagation algorithm (in the best case) was 33.

It is also possible to see in Table 2 that the minimum number of epochs of backpropagation is less than the number of the interactions of the quantum walk procedure, see Table 3. Although it appears that the classical backpropagation algorithm has an advantage over the quantum walk procedure, it is essential to observe that the number of interactions t is known before start the search in the quantum walk algorithm, but it is not for the classical backpropagation. Thus, the critical measure to characterize the practical cost expectation is the mean number of epochs for the classical backpropagation. Prior knowledge of the amount of interaction required for the algorithm to converge is a great advantage of the quantum walk algorithm.

Therefore, observing the mean result of epochs required for network training by the backpropagation (Table 2) and the values of t in Table 3, the computational proposed procedure is more efficient on average. Depending on the size of the search space of the proposed algorithm, the network training using the backpropagation algorithm, in some cases, performs a smaller number of iterations. However, on average, the iterations number of the proposed algorithm is decidedly smaller than the backpropagation procedure. As seen in the backpropagation simulations, the random initialization of weights is one factor that influences the result of time convergence for the network. However, because of the impossibility of determination for an excellent region to initialize the weights, the mean value and standard deviation are the information statistically relevant for backpropagation experiments, where low performance and high variation in training epochs number results are obtained when compared with the proposed quantum algorithm.

Nevertheless, the proposed algorithm is quantum. Table 4 shows the measurement probabilities for the five simulations for neural network training. It was considered both states $|AA\rangle$ and $|AB\rangle$ are solutions, given that for these states, the walker is in a vertex a . At least 98.55 percent of the

TABLE 4
Measurement Experiment — The Percentage Measure
Observed for Each Quantum

Experiment	$ AA\rangle$	$ AB\rangle$	$ BA\rangle$	$ BB\rangle$
1	95.48%	3.07%	1.40%	0.05%
2	95.03%	3.67%	1.26%	0.04%
3	100.0%	0.00%	0.00%	0.00%
4	99.99%	0.00%	0.01%	0.00%
5	99.88%	0.10%	0.02%	0.00%

The states $|AA\rangle$ and $|AB\rangle$ are the solutions.

measurements find a solution (experiment 1), reaching 100 percent for the experiment 3. At mean, the proposed quantum procedure find a solution in 99.44 percent of the measurements.

6 CONCLUSION

Training based on the backpropagation algorithm (or descendant gradient algorithms) may fall to local minimums. Many factors can influence this result and carry training for many epochs until the network converges to a solution or even stagnates. Therefore, it is impossible to say whether the training will stop or that good accuracy can be guaranteed in a prior way.

The non-stagnating in local minimums and the knowing in advance the number of iterations required to obtain a solution are some advantages of the proposed procedure. It is not possible to guarantee that the solution obtained is optimal because the probability is equal between all weight set solutions within the state $|AA\rangle$ and the state $|AB\rangle$. However, with a high probability, there will have a valid solution at the end of the procedure. So, the procedure proposed here guarantees the neural network training with high probability. However, the proposed methodology is a theoretical proposal. In practice, the proposed algorithm needs a quantum computer, which does not exist yet.

Another critical point is the oracle. In a quantum system, an oracle is an operator that can answer if a given state is or is not marked, i.e., if a given state is or is not a solution. For the proposed methodology, an oracle would be an operator capable of determining whether a given state would train a network or not. Here, an oracle was simulated by a simple matrix of zeros and ones. Each position in the oracle matrix is a possible state in the quantum system. If the value is 1, then the state is a solution. Otherwise, the state is not a solution. The creation of a real quantum oracle operator is a future research.

As seen before, the execution time of the proposed algorithm is of the order of $O(\sqrt{N/k})$, i.e., there a quadratic gain when compared with the classical analog. In this case, the search problem grows exponentially, the window's search space is $O(P^\omega)$, where $P = \sqrt{N}$, N is the number of points in the search window (a squared window), and ω is the number of ANN weights. The number of qubits required to represent ω synaptic weights is equal to $\omega * \log_2(N)$. The proof of correctness of the quantum walk algorithm can be found in [27].

Finally, classical simulations showed a significant gain in training an artificial neural network using the procedure

that applies a quantum walk to find the set of weights compared with the use of the backpropagation algorithm.

ACKNOWLEDGMENTS

The authors would like to thank the Science and Technology Support Foundation of Pernambuco (FACEPE) Brazil, the Brazilian National Council for Scientific and Technological Development (CNPq), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 by financial support for the development of this research.

REFERENCES

- [1] J. R. Powell, "The quantum limit to Moore's law," *Proc. IEEE*, vol. 96, no. 8, pp. 1247–1248, Aug. 2008.
- [2] J. Singh and M. Singh, "Evolution in quantum computing," in *Proc. Int. Conf. Syst. Model. Advancement Res. Trends*, 2016, pp. 267–270.
- [3] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning," 2013, *arXiv: 1307.0411*.
- [4] N. S. Yanofsky, M. A. Mannucci, and M. A. Mannucci, *Quantum Computing for Computer Scientists*, vol. 20. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [5] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proc. 28th Annu. ACM Symp. Theory Comput.*, 1996, pp. 212–219.
- [6] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. New York, NY, USA: Academic, 2014.
- [7] V. Dunjko, J. M. Taylor, and H. J. Briegel, "Quantum-enhanced machine learning," *Phys. Rev. Lett.*, vol. 117, no. 13, 2016, Art. no. 130501.
- [8] E. Aïmeur, G. Brassard, and S. Gambs, "Quantum speed-up for unsupervised learning," *Mach. Learn.*, vol. 90, no. 2, pp. 261–287, 2013.
- [9] Y. Zheng, S. Lu, and R.-B. Wu, "Quantum circuit design for training perceptron models," 2018, *arXiv: 1802.05428*.
- [10] M. Schuld, I. Sinayskiy, and F. Petruccione, "Simulating a perceptron on a quantum computer," *Phys. Lett. A*, vol. 379, no. 7, pp. 660–663, 2015.
- [11] P. Date, C. Schuman, R. Patton, and T. Potok, "A classical-quantum hybrid approach for unsupervised probabilistic machine learning," in *Proc. Future Inf. Commun. Conf.*, 2019, pp. 98–117.
- [12] E. Aïmeur, G. Brassard, and S. Gambs, "Machine learning in a quantum world," in *Proc. Conf. Can. Soc. Comput. Stud. Intell.*, 2006, pp. 431–442.
- [13] A. Prieto et al., "Neural networks: An overview of early research, current frameworks and new challenges," *Neurocomputing*, vol. 214, pp. 242–268, 2016.
- [14] S. Haykin, "Neural networks: Principles and practice," *Bookman*, vol. 11, 2001, Art. no. 900.
- [15] V. K. Ojha, A. Abraham, and V. Snášel, "Metaheuristic design of feedforward neural networks: A review of two decades of research," *Eng. Appl. Artif. Intell.*, vol. 60, pp. 97–116, 2017.
- [16] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, 2017, Art. no. 195.
- [17] T. G. Wong, "Grover search with lackadaisical quantum walks," *J. Phys. A: Math. Theor.*, vol. 48, no. 43, 2015, Art. no. 435304.
- [18] N. Nahimovs and A. Rivosh, "Exceptional configurations of quantum walks with Grover's coin," in *Proc. Int. Doctoral Workshop Math. Eng. Methods Comput. Sci.*, 2015, pp. 79–92.
- [19] T. G. Wong, "Faster search by lackadaisical quantum walk," *Quantum Inf. Process.*, vol. 17, no. 3, 2018, Art. no. 68.
- [20] N. B. Lovett, M. Everitt, R. M. Heath, and V. Kendon, "The quantum walk search algorithm: Factors affecting efficiency," *Math. Struct. Comput. Sci.*, vol. 29, no. 3, pp. 389–429, 2019.
- [21] N. Nahimovs, "Lackadaisical quantum walks with multiple marked vertices," in *Proc. Int. Conf. Current Trends Theory Practice Informat.*, 2019, pp. 368–378.
- [22] L. S. de Souza, J. H. A. de Carvalho, and T. A. E. Ferreira, "Quantum walk to train a classical artificial neural network," in *Proc. 8th Brazilian Conf. Intell. Syst.*, 2019, pp. 836–841.

- [23] R. Portugal, *Quantum Walks and Search Algorithms*. Cham, Switzerland: Springer, 2013.
- [24] S. E. Venegas-Andraca, "Quantum walks: A comprehensive review," *Quantum Inf. Process.*, vol. 11, no. 5, pp. 1015–1106, 2012.
- [25] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge University Press, 2000.
- [26] N. Shenvi, J. Kempe, and K. B. Whaley, "Quantum random-walk search algorithm," *Phys. Rev. A*, vol. 67, no. 5, 2003, Art. no. 052307.
- [27] Y. Feng, R. Duan, Z. Ji, and M. Ying, "Proof rules for the correctness of quantum programs," *Theor. Comput. Sci.*, vol. 386, no. 1/2, pp. 151–166, 2007.
- [28] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, "Quantum amplitude amplification and estimation," *Contemp. Math.*, vol. 305, pp. 53–74, 2002.
- [29] M. Boyer, G. Brassard, P. Høyer, and A. Tapp, "Tight bounds on quantum searching," *Fortschritte der Physik: Progress Phys.*, vol. 46, no. 4/5, pp. 493–505, 1998.
- [30] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015, Art. no. 436.
- [31] I. N. Da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, *Artificial Neural Networks: A Practical Course*. Cham, Switzerland: Springer, 2017.
- [32] Y. Sasaki and M. Nakahara, *Quantum Information and Quantum Computing*. Singapore: World Scientific, 2013.
- [33] A. Kapoor, N. Wiebe, and K. Svore, "Quantum perceptron models," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3999–4007.



Luciano S. de Souza received the BS degree in computer science from the Federal Rural University of Pernambuco, Recife, Brazil, in 2016, and the MS degree in applied informatics from the Federal Rural University of Pernambuco, Recife, Brazil, in 2019. He is currently working toward the PhD degree at the Federal Rural University of Pernambuco, Recife, Brazil. His research interests include quantum computing and machine learning.



Jonathan H. A. de Carvalho received the BS degree in information systems from the Federal Rural University of Pernambuco, Recife, Brazil, in 2019. Currently, he is working toward the MS degree in computer science with the Center of Informatics, Federal University of Pernambuco, Recife, Brazil. His research interests include quantum computing, machine learning, data mining, and spatial search.



Tiago A. E. Ferreira received the BS and MS degrees in physics from the Federal University of Pernambuco (UFPE), Recife, Brazil, in 1995 and 1998, respectively, the PhD degree in physics from São Paulo University, São Paulo, Brazil, in 2001, and the PhD degree in computer science from the UFPE, Recife, Brazil, in 2006. He is currently a professor with the Statistical and Informatics Department, Federal Rural University of Pernambuco and a visiting professor with Harvard University – Institute for Applied Computational Science. He is a member of the graduate programs in applied informatics, and biometrics and applied statistics. He is a member of the Brazilian Physical Society and of the Brazilian Computer Science Society. He is a researcher of the Brazilian National Research Council. His research interests include simulation and non-linear modeling, time series analysis and forecasting, machine learning, high-performance GP/GPU programming, quantum computing, and astroinformatics.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.