

A Quantum-Inspired Evolutionary Learning Process to Design Dilation-Erosion Perceptrons for Financial Forecasting

Ricardo de A. Araújo^{1,2}, Adriano L. I. Oliveira¹, Sergio Soares¹, Silvio Meira¹

¹Informatics Center, Federal University of Pernambuco, Recife, PE, Brazil

²Informatics Department, Federal Institute of Sertão Pernambucano, Ouricuri, PE, Brazil

ricardo.araujo@ifsertao-pe.edu.br or {raa,alio,scbs,srlm}@cin.ufpe.br

Abstract – Financial forecasting problems are rather difficult to be solved due to many complex features present in these time series. Several techniques have been proposed in the literature to solve this kind of problem. However, a dilemma arises from them, known as random walk dilemma, where the forecasts generated show a characteristic one step delay with respect to the real time series data. In this sense, this work presents a quantum-inspired evolutionary learning process to design the dilation-erosion perceptron (DEP) in order to overcome the random walk dilemma for financial forecasting. Furthermore, an experimental analysis is presented using the Dow Jones Industrial Average Index, where five well-known performance metrics and an evaluation function are used to assess forecasting performance.

Keywords: Dilation-Erosion Perceptron, Quantum-Inspired Evolutionary Learning, Financial Time Series Forecasting, Random Walk Dilemma.

1 INTRODUCTION

Financial forecasting problems are rather difficult to be solved due to many complex features frequently present in these time series, such as irregularities, volatility, trends and noise. Many efforts have been made to the development of linear and nonlinear statistical models able to determine the future behavior of financial phenomena [1–5].

Several alternative approaches have been proposed in the literature to solve this problem [6–13]. In the last two decades, the most popular approach for nonlinear modeling of time series is based on artificial neural networks (ANNs) [14, 15]. However, to define a solution to a given problem, ANNs require setting several system parameters, some of which are not always easy to determine. In this context, evolutionary approaches for the definition of neural network parameters have produced interesting results [16–23].

However, even with robust techniques to solve forecasting problems, a dilemma arises from all these models with respect to financial time series, known as random walk dilemma (RWD) [7, 24], where it is possible to verify that the forecasts generated by arbitrary models present a characteristic one step delay regarding real time series data [13, 21–23]. Therefore, as overcoming the RWD is a too hard task, some researchers have been argued that these time series cannot be predicted [7, 24].

In this sense, this work presents a quantum-inspired evolutionary learning process (using a quantum-inspired evolutionary algorithm (QIEA) [25]) to design the dilation-erosion perceptron (DEP) [13] in order to overcome the RWD for financial forecasting. Also, we have included an automatic phase fix procedure (APFP) [13, 21–23, 26–28] into proposed learning process in the attempt to adjust time phase distortions that occur in some temporal phenomena. Furthermore, an experimental analysis is conducted with the proposed model using the Dow Jones Industrial Average Index (DJI series), where we can demonstrate that the proposed model can successfully overcome the RWD, having good forecasting performance according to five well-known performance metrics and an evaluation function defined in [13].

2 FUNDAMENTALS

In this section we present the fundamentals and theoretical concepts for the proposed model.

2.1 THE TIME SERIES FORECASTING PROBLEM

A time series is a sequence of observations about a given phenomenon observed in a discrete or continuous space. In this work all time series will be considered time discrete and equidistant, and formally defined by

$$\mathbf{x} = \{x_t \in \mathbb{R} \mid t = 1, 2, \dots, N\}, \quad (1)$$

where t is the temporal index, which is called time and defines the granularity of observations of a given phenomenon, and N is the number of observations.

The aim of prediction techniques applied to a given time series (\mathbf{x}) is to provide a mechanism that allows, with certain accuracy, the prediction of the future values of x_t , given by x_{t+h} , $h = 1, 2, \dots, H$, where h represents the prediction horizon of H steps ahead. These prediction techniques try to identify certain regular patterns present in the data set, creating a model capable

of generating the next temporal patterns, where, in this context, a most relevant factor for an accurate prediction performance is the correct choice of the past window, or the time lags, considered for the representation of a given time series.

Box and Jenkins [1] shown that when there is a clear linear relationship among the historical data of a given time series, the functions of autocorrelation and partial autocorrelation are capable of identifying the relevant time lags to represent a time series, and such procedure is usually applied in linear models. However, when it uses real world time series, or more specifically, complex time series with all their dependencies on exogenous and uncontrollable variables, the relationship that involves the time series historical data is generally nonlinear, which makes the Box and Jenkins analysis procedure of the time lags only a crude estimate.

In mathematical sense, such relationship involving time series historical data defines a d -dimensional time phase space, where d is the minimum dimension capable of representing such relationship. Therefore, a d -dimensional time phase space can be built so that it is possible to unfold its corresponding time series. Takens [29] proved that if d is sufficiently large, such time phase space is homeomorphic to the time phase space that generates the series. Takens' Theorem [29] is the theoretical justification that it is possible to rebuild a phase space using the correct time lags. If this time phase space is correctly rebuilt, Takens' Theorem [29] also guarantees that the dynamics of this time phase space is topologically identical to the dynamics of the real system time phase space.

The main problem in reconstructing the original time phase space is naturally the correct choice of the variable d , or more specifically, the correct choice of the important time lags necessary for the characterization of the system dynamics. Many proposed methods can be found in the literature for the definition of the time lags [30–32]. These methods are usually based on measures of conditional probabilities, which consider,

$$x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-d}) + r_t, \quad (2)$$

where $f(x_{t-1}, x_{t-2}, \dots, x_{t-d})$ is a possible mapping of the pasts values to the facts of the future (where x_{t-1} is the lag 1, x_{t-2} is the lag 2, \dots , x_{t-d} is the lag d) and r_t is a noise term.

However, in general way, these tests found in the literature are based on the primary dependence among the variables and do not consider any possible induced dependencies. For example, if

$$f(x_{t-1}) = f(f(x_{t-2})), \quad (3)$$

then x_{t-1} is the primary dependence, and the dependence induced on x_{t-2} is not considered (any variable that is not a primary dependence is denoted as irrelevant). The model proposed in this work, conversely, does not make any prior assumption about the dependencies between the variables. In other words, it does not discard any possible correlation that can exist among the time series parameters, even higher order correlations.

2.2 THE RANDOM WALK DILEMMA

A naive prediction strategy is to define the last observation of a time series as the best prediction of its next future value ($x_{t+1} = x_t$). This kind of model is known as the random walk (RW) model [7, 33], which is defined by

$$x_t = x_{t-1} + r_t, \quad (4)$$

or

$$\Delta x_t = x_t - x_{t-1} = r_t, \quad (5)$$

where x_t is the current observation, x_{t-1} is the immediate observation before x_t , and r_t is a noise term with a Gaussian distribution of zero mean and standard deviation σ ($r_t \approx N(0, \sigma)$). In other words, the rate of time series change (Δx_t) is a white noise.

The model above clearly implies that, as the information set consists of past time series data, the future data is unpredictable. On average, the value x_t is indeed the best prediction of value x_{t-1} . This behavior is common in the finance and economics and is called random walk dilemma or random walk hypothesis [7, 33].

The computational cost for time series prediction using the random walk dilemma is extremely low. Therefore, any other prediction method more costly than a random walk model should have a very superior performance than a random walk model. Otherwise its use is not interesting in practice. However, if the time series phenomenon is driven by a law with strong similarity to a random walk model, any model applied to this phenomenon will tend to have the same performance as a random walk model.

Assuming that an accurate prediction model is used to build an estimated value of x_t , denoted by \hat{x}_t , the expected value ($E[\cdot]$) of the difference between \hat{x}_t and x_t must tend to zero,

$$E[\hat{x}_t - x_t] \rightarrow 0. \quad (6)$$

If the time series generator phenomenon is supposed to have a strong random walk linear component and a very weak nonlinear component (denoted by $g(t)$), and assuming that $E[r_t] = 0$ and $E[r_i r_j] = 0$ ($\forall i \neq j$), the expected value of the difference between \hat{x}_t and x_t (assuming that $x_t = x_{t-1} + g(t) + r_t$) will be

$$E[\hat{x}_t - (x_{t-1} + g(t) + r_t)] \rightarrow 0, \quad (7)$$

$$E[\hat{x}_t] - E[x_{t-1}] - E[g(t)] - E[r_t] \rightarrow 0, \quad (8)$$

$$E[\hat{x}_t] - E[x_{t-1}] - E[g(t)] \rightarrow 0, \quad (9)$$

$$E[\hat{x}_t] \rightarrow E[x_{t-1}] + E[g(t)]. \quad (10)$$

As nonlinear component of some time series (generally financial like) have small magnitude regarding linear component, it can assume that

$$E[g(t)] \rightarrow 0, \quad (11)$$

and consequently

$$E[x_{t-1}] + E[g(t)] \rightarrow E[x_{t-1}]. \quad (12)$$

Then, the equation 10 can be rewritten by

$$E[\hat{x}_t] \rightarrow E[x_{t-1}]. \quad (13)$$

It is possible to verify that the use of an arbitrary model to make forecasts have an intrinsic limitation, since the generated forecasts have a characteristic one step ahead delay regarding the original time series values, in which this behavior is common in the finance and economics and is called random walk dilemma or random walk hypothesis [7]. Therefore, in these conditions, to escape of the random walk dilemma is a hard task [13].

3 THE DILATION-EROSION PERCEPTRON (DEP)

According to Araújo [13], financial forecasting problems can be modeled in terms of functions $\Psi : \mathbb{R}_{\pm\infty}^d \rightarrow \mathbb{R}_{\pm\infty}$ (d represents the minimum necessary dimension to determining the characteristic phase space that generates the time series phenomenon, or, the time lags dimensionality), which can be approximated in terms of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, and given by

$$\Psi \simeq \delta_{\mathbf{a}} \quad \text{and} \quad \Psi \simeq \varepsilon_{\mathbf{b}}, \quad (14)$$

where

$$\delta_{\mathbf{a}}(\mathbf{x}) = \bigvee_{i=1}^d (x_i + a_i) \quad \text{and} \quad \varepsilon_{\mathbf{b}}(\mathbf{x}) = \bigwedge_{i=1}^d (x_i + b_i), \quad (15)$$

in which $\mathbf{x} \in \mathbb{R}^d$, terms \bigvee and \bigwedge represent infimum and supremum operators [13], and the main differences between “+” and “+’” are given by the following rules:

$$(-\infty) + (+\infty) = (+\infty) + (-\infty) = -\infty, \quad (16)$$

and

$$(-\infty) + ' (+\infty) = (+\infty) + ' (-\infty) = +\infty. \quad (17)$$

Let $\mathbf{x} \in \mathbb{R}^d$ a real-valued input signal inside an d -point moving window of the time series and let y the output of the DEP. Then, the DEP is defined by a translation invariant morphological operator (Ψ like) with local signal transformation rule $\mathbf{x} \rightarrow y$, given by

$$y = \lambda\alpha + (1 - \lambda)\beta, \quad \lambda \in [0, 1], \quad (18)$$

with

$$\alpha = \delta_{\mathbf{a}}(\mathbf{x}), \quad (19)$$

and

$$\beta = \varepsilon_{\mathbf{b}}(\mathbf{x}), \quad (20)$$

where $\lambda \in \mathbb{R}$, terms $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ represent the structuring elements of morphological operators of dilation and erosion, respectively. In this way, it is worth mentioning that the DEP have a convex combination of its components, where when it increases the contribution of one component, the other one tends to decrease. The structure of DEP is illustrated in Figure 1.

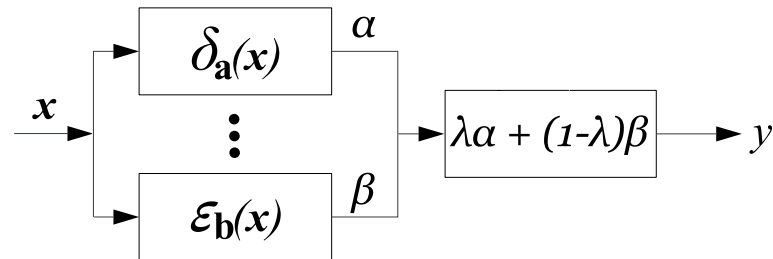


Figure 1: The proposed DEP structure.

4 THE PROPOSED QUANTUM-INSPIRED EVOLUTIONARY LEARNING PROCESS

According to the DEP definition, we can see that the main objective of its design is to determine a set of parameters defined by $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^d$ and λ . Therefore, the weight vector ($\mathbf{w} \in \mathbb{R}^n$ with $n = 2d + 1$) to be used in the learning process is given by

$$\mathbf{w} = (\mathbf{a}, \mathbf{b}, \lambda). \quad (21)$$

During the proposed quantum-inspired evolutionary learning process, the weights of the DEP are adjusted according to an error criterion until convergence or until the end of quantum-inspired search generations. Each i -th individual from classical population at g -th generation represents a candidate weight vector (denoted by $\mathbf{w}_i^{(g)}$) for the DEP model. The scheme to adjust the weight vector is initially to define a fitness function f (which must reflect the solution quality achieved by the parameters configuration of the system), given by

$$f(\mathbf{w}_i^{(g)}) = \frac{1}{M} \sum_{j=1}^M e^2(j), \quad (22)$$

where M is the number of input patterns and $e(j)$ is the instantaneous error, given by

$$e(j) = t(j) - y(j), \quad (23)$$

where $t(j)$ and $y(j)$ are the target output and the actual model output for the j -th training pattern, respectively.

The proposed quantum-inspired evolutionary learning process, called DEP(QIEA), employ a quantum-inspired evolutionary algorithm (QIEA) to train the DEP model. According to previous experiments with some quantum-inspired evolutionary algorithms versions [25, 34–36], we decided to use the version presented in [25], which employ a novel real-valued representation for quantum individuals in order to explore the state space more efficiently and to enhance convergence speed. A further discussion about this choice is beyond the scope of this work, and we refer the reader to an upcoming journal paper.

The QIEA procedure performs some steps to minimize the fitness function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, which is defined by Equation 22. Recall that term n represents the dimensionality of the DEP model weight vector, which is given by $2d + 1$. These steps consists on the generation of quantum individuals, using the concept of quantum bits (qubits) and the superposition of states, to build a set of classical individuals, using the interference process. At the end, the best individual in the classical population is selected as a solution to the problem. In our simulations, both quantum and classical populations comprises ten individuals ($S = 10$).

The first step is to build the quantum population, which is given by a superposition of states that are observed to generate classical individuals (candidate solutions of the problem), and defined by

$$\mathbf{QP}^{(g)} = (\mathbf{QP}_1^{(g)}, \mathbf{QP}_2^{(g)}, \dots, \mathbf{QP}_S^{(g)}), \quad (24)$$

with

$$\mathbf{QP}_i^{(g)} = (QP_{i1}, QP_{i2}, \dots, QP_{in}), \quad (25)$$

and

$$QP_{ij} = (\rho_{ij}, \sigma_{ij}), \quad (26)$$

in which $\mathbf{QP}^{(g)}$ denotes the quantum population at generation g ; $\mathbf{QP}_i^{(g)}$ denotes i -th quantum individual of the population $\mathbf{QP}^{(g)}$; QP_{ij} denotes the j -th parameter of the i -th individual of the population. Terms $\rho_{ij}, \sigma_{ij} \in \mathbb{R}$ represents the center and the width of a square pulse, which is used to build the set of possible observable values over the problem domain [25]. The height (h_{ij}) of each pulse is defined using the quantum gene width (σ_{ij}) and the maximum number of quantum individuals (S) in quantum population, given by [25]

$$h_{ij} = \frac{1/\sigma_{ij}}{S}. \quad (27)$$

The second step is to build the classical population, where we use the interference process among quantum individuals to generate a probability density function (PDF). The PDF consists of summing up the quantum individuals genes, that is, the first gene of all quantum individuals are summed, and all other genes of a given quantum individual do the same. The PDF is defined by [25]

$$PDF_j = \sum_{i=1}^S QP_{ij}, \quad (28)$$

where QP_{ij} represents the square pulse with width σ_{ij} and center ρ_{ij} of the j -th gene of the i -th quantum individual.

These PDFs are used to build the classical individuals, which are real-valued vectors with same amount of quantum individuals genes, where these values are randomly selected using the PDFs as probability function. In the attempt to perform a random selection, it is necessary to define a cumulative distribute function (CDF), which is given by [25]

$$CDF_j(x) = \int_l^u PDF_j(x) dx, \quad (29)$$

where u and l represent the upper and lower bounds of PDF_j function.

Then, as all PDFs are built by a sum of square pulses, we can calculate the PDF area by dividing the function curve in rectangles and by summing up its corresponding area. Note that CDFs can be calculated using these PDFs based on such rectangles [25]. Through these CDFs, we can build a set of classical individuals by using such curves. Therefore, the classical population is created by an uniform choice of random numbers in the range $[0, 1]$ and by the identification of these points in CDF. Therefore, we can generate the j -th parameter of i -th individual at g generation of classical population by [25]

$$w_{ij}^{(g)} = CDF^{-1}(r), \quad (30)$$

where r is a random number in the range $[0, 1]$.

This procedure allows to build the temporary classical population (TCP), which stores all classical individuals generated using the quantum population. At first QIEA generation, the classical population (CP), which is the best observations (in terms of fitness function) of quantum population, is a clone of the TCP . For next generations, the multi-point crossover operator [25] is applied in the classical QIEA population to generate better classical individuals, hence improving the quantum population update process.

After the CP generation, it is necessary to update the quantum population. First we use a translate operation, which is responsible to update the center (ρ) of each quantum genes. A simple procedure to do this is to replace the mean of each gene values to the genes values from classical individuals. This step is formally defined by

$$\rho_{ij} = w_{ij}^{(g)}, \quad (31)$$

where ρ_{ij} represents the center of j -th gene of the i -th quantum individual from quantum population, and $w_{ij}^{(g)}$ denotes the j -th gene of the i -th classical individual at g generation from CP .

Then we use a resize operation, which is responsible to reducing or enlarging the width (σ_{ij}) of quantum gene. This change should be made homogeneously for all quantum genes and for all quantum individuals. We use the $1/5$ th rule to determine if such width should be enlarged or reduced, which is given by [25]

$$\sigma_{ij} = \begin{cases} \sigma_{ij} \cdot r & \text{if } \varphi < 1/5 \\ \sigma_{ij}/r & \text{if } \varphi > 1/5 \\ \sigma_{ij} & \text{otherwise} \end{cases}, \quad (32)$$

where σ_{ij} represents the width of j -th gene of the i -th quantum individual from quantum population, r denotes a random number in interval $[0, 1]$, and φ is the rate of how many classical individuals generated in a new generation have their overall evaluation improved.

Besides, in order to automatically adjust time phase distortions in some time series representation, we have included an automatic phase fix procedure (APFP) [13] in the proposed learning process of the DEP model. Figure 2 presents the APFP.

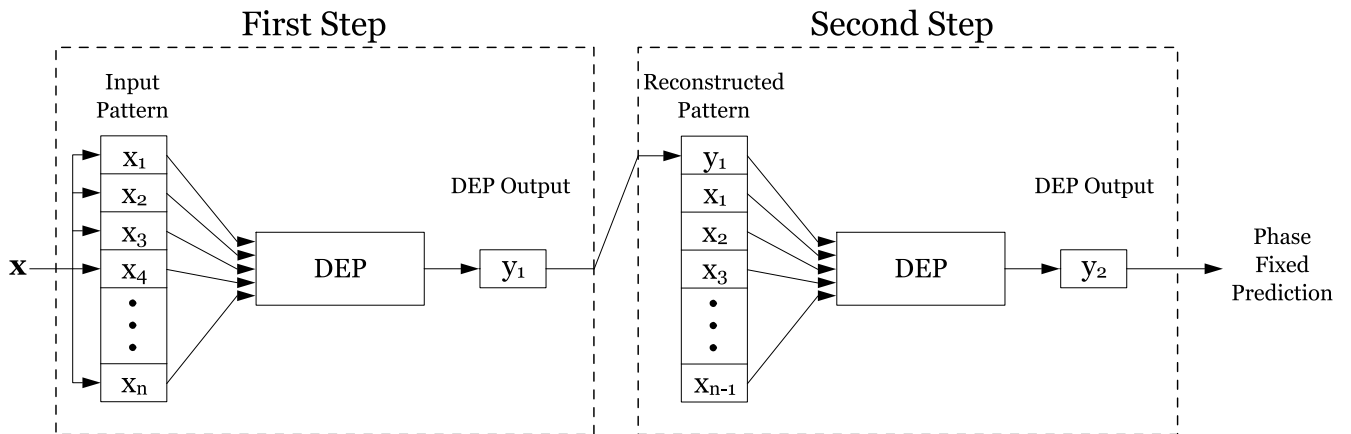


Figure 2: Automatic phase fix procedure.

According to Figure 2, in the first step an input pattern x is presented to DEP generating the output y_1 . The first output y_1 is used to rebuild the input pattern in the second step. This reconstructed pattern is presented to the same DEP generating the second output y_2 , which is the phase fixed forecasting.

Figure 3 presents the proposed quantum-inspired evolutionary learning process steps including the APFP.

It is worth mentioning that three stop conditions are used in the proposed learning process:

1. The maximum generation number: $g = 10^4$;
2. The decrease in the training error process training (Pt) [37] of the cost function: $Pt \leq 10^{-6}$.
3. The increase in the validation error or generalization loss (Gl) [37] of the cost function: $Gl > 5\%$;

```

begin DEP Learning Process
   $g = 0;$  //  $g$ : actual generation
  create quantum population;
  initialize the stop condition;
  while not stop condition do
     $g = g + 1;$ 
    create the PDFs using quantum individuals;
    for  $i = 1$  to  $S$  do
      create the temporary classical individual  $w_i^{(g)}$  observing quantum population and using CDFs;
      initialize DEP parameters with the values supplied by  $w_i^{(g)}$ ;
      calculate  $y_1, y_2$  and the instantaneous error for all input patterns;
      evaluate the temporary classical individual  $f(w_i^{(g)})$  using the Equation 22;
    end
    if  $g=1$  then
      | classical population = temporary classical population;
    else
      | temporary classical population = crossover operator between current temporary classical population
      | and classical population;
      initialize DEP parameters with the values supplied by temporary classical population;
      calculate  $y_1, y_2$  and the instantaneous error for all input patterns;
      evaluate temporary classical population;
      | classical population =  $K$  best individuals from temporary classical population;
    end
    apply translate operation;
    apply resize operation;
  end
end

```

Figure 3: Quantum-inspired evolutionary learning steps.

5 EXPERIMENTAL RESULTS

The Dow Jones Industrial Average Index (DJI time series) was used as a test bed for evaluation of the proposed model. The time series was normalized to lie within the range $[0, 1]$ and divided in three sets according to Prechelt [37]: training set (50% of the data points), validation set (25% of the data points) and test set (25% of the data points). All constant parameters of the QIEA used in the proposed learning process are the same values suggested by Cruz [25].

In order to establish a performance study, results with the random walk (RW) model [24], which represents the results generated by classical forecasting models, is employed in our comparative analysis, where we investigate the same time series under the same conditions. Additionally, we have used five well-known evaluation metrics formally defined in [13] to assess the forecasting performance: mean square error (MSE), mean absolute percentage error (MAPE), u of theil statistic (UTS), prediction of change in direction (POCID) and average relative variance (ARV). Also, we use an evaluation function (EF) defined in [13] to serve as a global performance indicator for the proposed forecasting model. For each time series, five experiments were performed, where we calculate the mean (MEAN) and the standard deviation (STD) in the attempt to obtain an average forecasting performance of the proposed DEP(QIEA) model. Also, we calculate all confidence intervals (CI) with the assumption of normal distribution with 99% of certainty degree.

In addition, we include in our analysis an additional measure referred to as the percentage gain (PG), which measures, in percentage terms, how much better is the DEP(QIEA) regarding RW model. The PG is formally defined by

$$PG = 100 - 100 \frac{pmm}{imm}, \quad (33)$$

or

$$PG = 100 \frac{pmm}{imm} - 100, \quad (34)$$

in which pmm and imm represent the evaluation metric value found by DEP(QIEA) and by RW, respectively. Note that the Equation 33 must be used to measure the obtained gains for MSE, MAPE, UTS and ARV metrics, while the Equation 34 must be used to measure the obtained gains for POCID and EF metrics.

5.1 DJI SERIES

The Dow Jones Industrial Average Index is the main important international stock market index, which shows how thirty large, publicly-owned companies based in the United States have traded during a standard trading session in the stock market.

The DJI series corresponds to daily records of Dow Jones Industrial Average Index from 1998/01/01 to 2003/08/26. For the DJI series forecasting (with one step ahead of forecasting horizon – $H = 1$), we use the same time lags presented in [13] to create the input patterns (lags 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 – note that here $d = 10$). The Table 1 shows the experiments performed with the DEP(QIEA) model, where we calculate all evaluation metrics, as well as their MEAN, STD and CI.

Table 1: Results for all experiments with the proposed DEP(QIEA) model for DJI series (test set).

Statistics	Evaluation Metrics					
	MSE	MAPE	UTS	ARV	POCID	EF
	1.4527e-008	4.8224e-004	1.7309e-005	5.9471e-007	99.43	99.3837
	2.9588e-007	2.1764e-003	3.5256e-004	1.2113e-005	99.43	99.1814
	2.7202e-008	6.5990e-004	3.2413e-005	1.1136e-006	99.43	99.3645
	2.3716e-006	6.1618e-003	2.8260e-003	9.7095e-005	99.43	98.5380
	7.3588e-007	3.4323e-003	8.7685e-004	3.0127e-005	99.43	99.0038
MEAN	6.8903e-007	2.5825e-003	8.2102e-004	2.8209e-005	99.43	99.0943
STD	9.8501e-007	2.3348e-003	1.1737e-003	4.0326e-005	0.00	0.3471
CI	$\pm 1.1365e-006$	$\pm 2.6940e-003$	$\pm 1.3542e-003$	$\pm 4.6528e-005$	± 0.00	± 0.4005

According to the Table 1, we can see that in all experiments the POCID metric greater than 50%, indicating that the DEP(QIEA) model has much better performance than a “coin-tossing” experiment. The obtained UTS metric value ($\simeq 8.2e-004$) indicates that the DEP(QIEA) model was able to overcome the random walk dilemma. Note that the MAPE metric value ($\simeq 2.6e-003$) is very small, that is, without high percentage deviations. According to ARV metric value ($\simeq 2.8e-005$), we can see a much better performance of the proposed model regarding a naive forecasting model. Also, we can verify a small value of MSE metric ($\simeq 6.9e-007$), which means that the forecasts are too close to real values. The EF metric value ($\simeq 99.4$) shows that the DEP(QIEA) have good global forecasting performance. We can also notice that the proposed model obtained small STD values, demonstrating the stability of the QIEA to train the DEP model.

Table 2 presents a performance study with the best results obtained by the proposed DEP(QIEA) with those presented with the RW model for DJI series.

Table 2: Best results (test set) for DJI series with RW and DEP(QIEA) models.

Evaluation Metrics	RW	DEP(QIEA)
MSE	8.3877e-004	1.4527e-008
MAPE	9.6687e-002	4.8224e-004
UTS	1.0000e-000	1.7309e-005
ARV	3.4338e-002	5.9471e-007
POCID	46.46	99.43
EF	21.7931	99.3837

Analyzing the Table 2, we can note that the proposed DEP(QIEA) model overcame the RW model in this work. However, to take more precise indications of the best performance of the proposed model, we present in Table 3 the obtained PG of the DJI series.

Table 3: Percentage gain (test set) for DJI series of the DEP(QIEA) regarding the RW.

Evaluation Metrics	PG (%) DEP(QIEA) / RW
MSE	100.00
MAPE	99.50
UTS	100.00
ARV	100.00
POCID	114.01
EF	356.03

According to Table 3 we can verify a better forecasting performance of the DEP(QIEA) regarding the RW (having a PG equals to 100% for all metrics, except for MAPE metric, having a PG around 99%). In addition, assessing the DEP(QIEA) in terms of overall forecasting performance (using EF metric), we have a PG around 356% regarding RW model.

Finally, we present in Figure 4 a comparative graphic between real (solid line) and predicted (dashed line) values generated by DEP(QIEA) and RW model for the last ten points of the DJI series test set. We can note that the predicted values are superimposed to the real values of the DJI series, where the one step delay regarding the forecasting values did not occur, that is, the time phase distortion that causes the random walk dilemma was successfully adjusted.

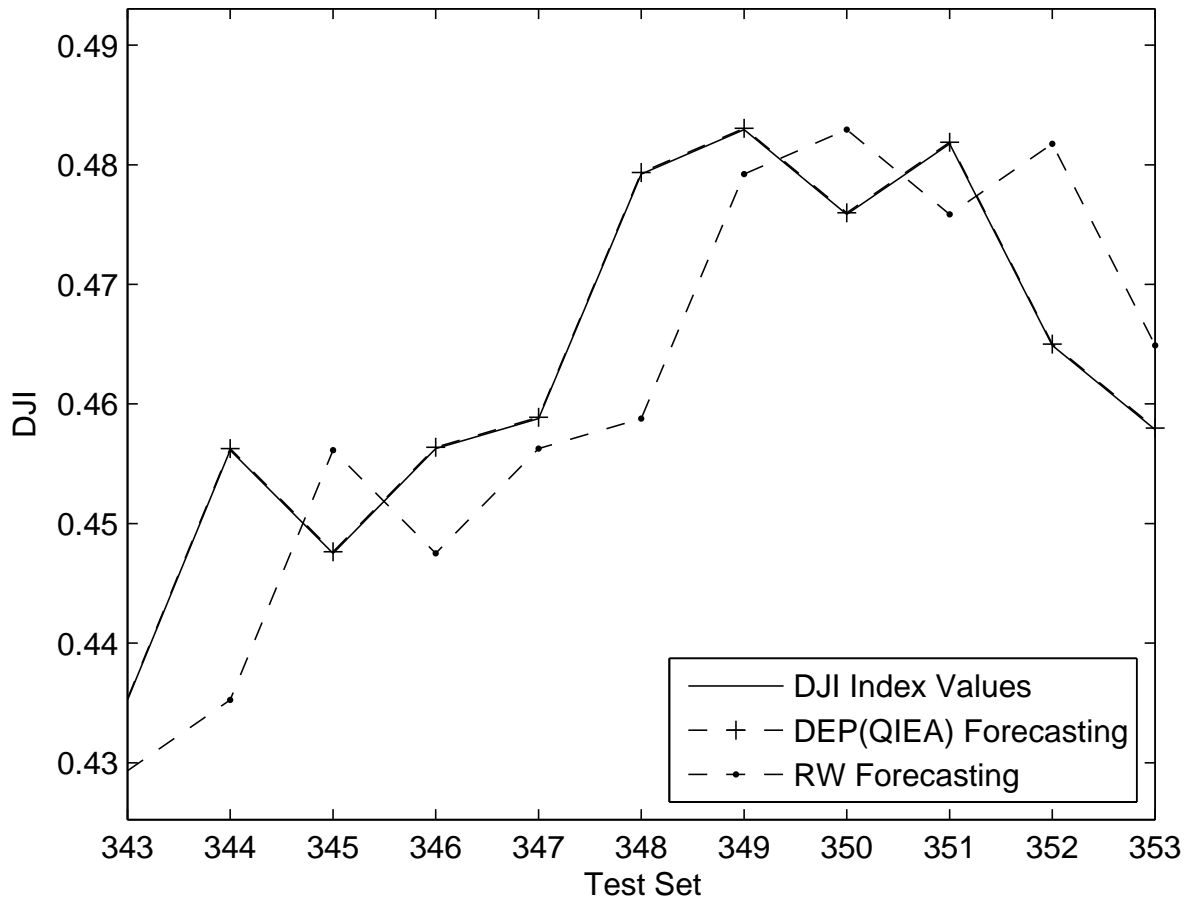


Figure 4: Forecasting results of DJI series (last ten points of the test set): actual values (solid line) and predicted values (dashed line).

6 CONCLUSION

In this work we presented a quantum-inspired learning process to design dilation-erosion perceptrons (DEP) to overcome the random walk dilemma for financial forecasting. The evaluation performance of the proposed DEP(QIEA) model regarding to random walk (RW) model was assessed in terms of five well-known performance measures and using the DJI series (with all their dependencies on exogenous and uncontrollable variables). In addition, an evaluation function served as a global indicator for the quality of solutions achieved by the investigated models.

The experimental results demonstrated a consistently better performance, of the proposed learning process, for training the DEP model. With the inclusion of the APFP into the proposed learning process of the DEP model, we succeeded in automatically correcting the time phase distortions that typically occur in financial forecasting problems, where our forecasts have not any one step delay regarding real time series values. A feasible explanation for such behavior is that the APFP depend on the information complexity contained in the time series data and the ability to accurately define the best forecasting model parameters to estimate the real time series values, in other words, the success of the APFP is strongly dependent on an accurate adjustment of the forecasting model parameters. Therefore, we can verify that the QIEA used to train the DEP model was able to adjust more precisely time phase distortions that occur in the analyzed DJI series.

Further studies must be developed to better formalize and explain the properties of the proposed model and to determine its possible limitations with other time series with components such as trends, seasonalities, impulses, steps and other nonlinearities.

ACKNOWLEDGEMENT

This work was partially supported by the National Institute of Science and Technology for Software Engineering (INES), funded by CNPq and FACEPE, grants 573964/2008-4 and APQ-1037-1.03/08.

References

- [1] G. E. P. Box, G. M. Jenkins and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, New Jersey, third edition, 1994.
- [2] T. S. Rao and M. M. Gabr. *Introduction to Bispectral Analysis and Bilinear Time Series Models*, volume 24 of *Lecture Notes in Statistics*. Springer, Berlin, 1984.
- [3] T. Ozaki. *Nonlinear Time Series Models and Dynamical Systems*, volume 5 of *HandBook of Statistics*. Noth-Holland, Amsterdam, 1985.
- [4] M. B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press, 1988.
- [5] D. E. Rumelhart and J. L. McClelland. *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*, volume 1 & 2. MIT Press, 1987.
- [6] T. V. Gestel, J. A. K. Suykens, D. E. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. D. Moor and J. Vandewalle. “Financial Time Series Prediction using Least Squares Support Vector Machines within the Evidence Framework”. *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 809–821, 2001.
- [7] R. Sitte and J. Sitte. “Neural Networks Approach to the Random Walk Dilemma of Financial Time Series”. *Applied Intelligence*, vol. 16, no. 3, pp. 163–171, May 2002.
- [8] M. P. Clements, P. H. Franses and N. R. Swanson. “Forecasting economic and financial time-series with Non-linear models”. *International Journal of Forecasting*, vol. 20, pp. 169–183, 2004.
- [9] P. Sussner and M. E. Valle. “Morphological and Certain Fuzzy Morphological Associative Memories for Classification and Prediction”. In *Computational Intelligence Based on Lattice Theory*, edited by V. G. Kaburlassos and G. X. Ritter, volume 67, pp. 149 – 173. Springer Verlag, Heidelberg, Germany, 2007.
- [10] Z. Shi and M. Han. “Support Vector Echo-State Machine for Chaotic Time Series Prediction”. *IEEE Transactions on Neural Networks*, vol. 18, no. 2, pp. 359–372, 2007.
- [11] R. de A. Araújo. “Swarm-based Hybrid Intelligent Forecasting Method for Financial Time Series Prediction”. *Learning and Nonlinear Models*, vol. 5, no. 2, pp. 137–154, 2007.
- [12] P. Sussner, R. Miyasaki and M. E. Valle. “An Introduction to Parameterized IFAM Models with Applications in Prediction”. In *Proceedings of IFSA-EUSFLAT 2009*, pp. 3024–3031, Lisbon, Portugal, 2009.
- [13] R. de A. Araújo. “A Class of Hybrid Morphological Perceptrons with Application in Time Series Forecasting”. *Knowledge-Based Systems*, 2011. . Accepted for Publication.
- [14] G. Zhang, B. E. Patuwo and M. Y. Hu. “Forecasting with Artificial Neural Networks: The State of the Art”. *International Journal of Forecasting*, vol. 14, pp. 35–62, 1998.
- [15] G. P. Zhang and D. M. Kline. “Quarterly Time-Series Forecasting With Neural Networks”. *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1800–1814, Nov. 2007.
- [16] O. Castillo and P. Melin. “Hybrid Intelligent Systems for Time Series Prediction using Neural Networks, Fuzzy Logic, and Fractal Theory”. *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1395–1408, 2002.
- [17] F. H. F. Leung, H. K. Lam, S. H. Ling and P. K. S. Tam. “Tuning of the Structure and Parameters of the Neural Network Using an Improved Genetic Algorithm”. *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 79–88, January 2003.
- [18] A. R. L. Junior, T. A. E. Ferreira and R. de A. Araújo. “An Experimental Study With a Hybrid Method for Tuning Neural Network for Time Series Prediction”. In *IEEE Congress on Evolutionary Computation*. IEEE, 2008.
- [19] T. A. E. Ferreira, G. C. Vasconcelos and P. J. L. Adeodato. “A New Intelligent System Methodology for Time Series Forecasting with Artificial Neural Networks”. In *Neural Processing Letters*, volume 28, pp. 113–129, 2008.
- [20] G. C. Onwubolu. “Design of hybrid differential evolution and group method of data handling networks for modeling and prediction”. *Information Sciences*, vol. 178, no. 18, pp. 3616–3634, 2008.
- [21] R. de A. Araújo. “Swarm-based Translation-invariant Morphological Method for Financial Time Series Forecasting”. *Information Sciences*, vol. 180, no. 24, pp. 4784–4805, 2010.
- [22] R. de A. Araújo. “Hybrid Intelligent Methodology to Design Translation Invariant Morphological Operators for Brazilian Stock Market Prediction”. *Neural Networks*, vol. 23, no. 10, pp. 1238–1251, 2010.

- [23] R. de A. Araújo. “Translation Invariant Morphological Time-lag Added Evolutionary Forecasting Method for Stock Market Prediction”. *Expert Systems with Applications*, vol. 38, no. 3, pp. 2835–2848, 2011.
- [24] B. G. Malkiel. *A Random Walk Down Wall Street, Completely Revised and Updated Edition*. W. W. Norton & Company, April 2003.
- [25] A. V. A. da Cruz, M. M. B. R. Velasco and M. A. C. Pacheco. “Quantum-Inspired Evolutionary Algorithm for Numerical Optimization”. In *Proceedings of the IEEE Congress on Evolutionary Computation*, Vancouver, Canada, 2006.
- [26] R. de A. Araújo and T. A. E. Ferreira. “An intelligent hybrid morphological-rank-linear method for financial time series prediction”. *Neurocomputing*, vol. 72, no. 10-12, pp. 2507–2524, 2009.
- [27] R. de A. Araújo and T. A. E. Ferreira. “A Morphological-Rank-Linear evolutionary method for stock market prediction”. *Information Sciences*, 2010. . In Press.
- [28] R. de A. Araújo. “A Hybrid Intelligent Morphological Approach for Stock Market Forecasting”. *Neural Processing Letters*, vol. 31, no. 3, pp. 195–217, 2010.
- [29] F. Takens. “Detecting Strange Attractor in Turbulence”. In *Dynamical Systems and Turbulence*, edited by A. Dold and B. Eckmann, volume 898 of *Lecture Notes in Mathematics*, pp. 366–381, New York, 1980. Springer-Verlag.
- [30] R. Savit and M. Green. “Time Series and Dependent Variables”. *Physica D*, vol. 50, pp. 95–116, 1991.
- [31] H. Pi and C. Peterson. “Finding the Embedding Dimension and Variable Dependences in Time Series”. *Neural Computation*, vol. 6, pp. 509–520, 1994.
- [32] N. Tanaka, H. Okamoto and M. Naito. “Estimating the Active Dimension of the Dynamics in a Time Series Based on a Information Criterion”. *Physica D*, vol. 158, pp. 19–31, 2001.
- [33] T. C. Mills. *The Econometric Modeling of Financial Time Series*. Cambridge University Press, Cambridge, 2003.
- [34] K. H. Han and J. H. Kim. “Quantum-inspired evolutionary algorithm for a class of combinatorial optimization”. *IEEE Trans. Evolutionary Computation*, vol. 6, no. 6, pp. 580–593, 2002.
- [35] J. S. Jang, K. H. Han and J. H. Kim. “Face Detection using Quantum-inspired Evolutionary Algorithm”. In *Proceedings of the IEEE Congress on Evolutionary Computation*, Portland, Oregon, 2004.
- [36] K. H. Han and J. H. Kim. “Quantum-inspired evolutionary algorithms with a new termination criterion, H_{ϵ} gate, and two-phase scheme.” *IEEE Trans. Evolutionary Computation*, vol. 8, no. 2, pp. 156–169, 2004.
- [37] L. Prechelt. “Proben1: A set of Neural Network Benchmark Problems and Benchmarking Rules”. Technical Report 21/94, 1994.