

# A Study for Multi-Objective Fitness Function for Time Series Forecasting with Intelligent Techniques

Aranildo R. L. Junior  
Department of Physics  
Federal University of Pernambuco  
Recife, Pernambuco, Brazil  
proj.brain@gmail.com

## ABSTRACT

This paper presents a study of a Hybrid method for time series prediction, called GRASPES, based on Greedy Randomized Adaptive Search Procedure (GRASP) Algorithm and Evolutionary Strategies (ES) concepts for tuning of the structure and parameters of an Artificial Neural Network (ANN). An experimental investigation with two time series is conducted and the results achieved are discussed and compared to other works reported in the literature. Distinct fitness functions evaluations are shown, instead of conventional MSE or NMSE based fitness functions evaluation. This results shown that small changes of the fitness function evaluation could lead to a significantly improved performance.

## Categories and Subject Descriptors

I.6.5 [Model Development]: Modeling methodologies; I.2.8 [Problem Solving, Control Methods, and Search]: Heuristic methods

## General Terms

Experimentation

## Keywords

Evolutionary Strategies, Neural Network, Time Series, Forecasting, Fitness Function

## 1. INTRODUCTION

Approaches based on artificial intelligence have been proposed for solve the nonlinear modeling of time series [1, 4], in particular Artificial Neural Networks (ANNs) has the ability to modeling the complex nonlinear relationships among data, without any prior assumptions about the nature of such relationships.

In this paper, a brief explanation about Time Series Forecasting Problem and Artificial Neural Networks (ANN), Greedy Randomized Adaptive Search Procedures (GRASP), Evolutionary Strategies (ES) is given and the systematic procedure based on this hybrid intelligent system is presented. The GRASPES, has the capability of automatically found

the parameters that require a previous definition to represent and solve the problem of time series forecasting. Then, the comparison of the experimental results with other methods and works of the literature and a systematic study for multi-objective fitness function is done, using two complex time series: Sunspot and Dow Jones Industrial Average (DJIA), where five performance measures were used (MSE, MAPE, NMSE, POCID and ARV).

## 2. FOUNDATIONS

### 2.1 The Time Series Forecasting Problem

A time series is a set of points, generally time equidistant, defined by,  $X_t = \{x_t \in \mathbb{R} \mid t = 1, 2, 3 \dots N\}$ , where  $t$  is the temporal index and  $N$  is the number of observations. Therefore,  $X_t$  is a sequence of temporal observations orderly sequenced and equally spaced.

The objective of the forecast problem is to apply some prediction techniques for the time series  $X_t$  and to identify patterns presents in the historical data, building a model able to identify the next time patterns.

Such relationship structures among historical data constitute a  $d$ -dimensional state space, where  $d$  is the minimum dimension capable of representing such relationship. Therefore, a  $d$ -dimensional state space can be built so that it is possible to unfold a time series in its interior. Takens [8] has proved that if  $d$  is sufficiently large, such built state space is homeomorphic to the state space which generated the time series. Thus, Takens Theorem [8] has provided the theoretical justification that it is possible to build a state space using the correct time lags, and if this space is correctly rebuilt, Takens Theorem [8] also guarantees that the dynamics of this space is topologically identical to the dynamics of the original system's state space.

As the intuit of this work is to predict continuous functions, and according to Cybenko [2], then will be used Multi-Layer Perceptron (MLP) networks with three layers of type  $i$ - $j$ - $k$ , where  $i$  denotes the number of time lags (input layer),  $j$  denotes the number of processing units in hidden layer (sigmoidal units) and  $k$  denotes the number of processing units in output layer ( $k = 1$  - prediction horizon is of one step ahead).

A sigmoidal activation function is used for all hidden processing units. The output processing unit used a linear activation function, where a sigmoidal function is applied to its bias. The ANN output is given by

$$y_k(t) = \sum_{j=1}^{n_h} \delta(S_{jk}^2) W_{jk} \text{Sig} \left[ \sum_{i=1}^{n_{in}} (\delta(S_{ij}^1) W_{ij} Z_i(t) - \delta(S_j^1) b_j^1) \right] - \delta(S_k^2) \text{Sig}(b_k^2), \quad (1)$$

where  $Z_i(t)$  ( $i = 1, 2, \dots, n_{in}$ ) are the ANN input values,  $n_{in}$

denotes the number of ANN input and  $n_h$  is the number of hidden units. Since the prediction horizon is one step ahead, only one output unit is necessary ( $k = 1$ ). The term  $Sig(\cdot)$  is a sigmoidal function defined by:

$$Sig(x) = \frac{1}{1 + exp(-x)}, \quad (2)$$

and  $\delta(\cdot)$  is a step function, defined by,

$$\delta(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

## 2.2 The GRASP Procedure

The GRASP (*Greedy Randomized Adaptive Search Procedures*) is a methodology that has a strong intuitive appeal, a register of prominent empirical track, and are trivial to execute. The GRASP is basically, a multistart procedure, where each iteration is made up of construction phase and randomized greedy solution is constructed. Then a local search phase which starts at the constructed solution and applies iterative improvement until a locally optimal solution is found of the considered problem [7].

A problem of combinatorial optimization, is defined by the finite set of data  $D = \{1, 2, \dots, N\}$ , the set of possible solutions  $G \subseteq 2^D$ , and the objective function  $f : 2^D \rightarrow \mathbf{R}$ . for minimization problems, searches for the excellent solution  $S' \in G$  such that  $f(S') = \min_{S \in G} f(S)$ . The ground set  $D$ , the cost function  $f$ , and the set of feasible solutions  $G$  are defined for each specific problem for example [7].

## 2.3 The Evolutionary Strategies

The nature ideas are applied to create the Evolutionary Algorithms (EA), which are composed by set of trial solutions of the problem (population), with each solution (individual) is coded by a parameter vector (data structure generically referred to as chromosome). Then, the genetic operations (crossover and mutation) are applied to these set of individuals in order to create the offspring (new generation). The offspring are similar to their parents, then each new generation has organisms that are similar to the fit members of the previous generation [6].

The Evolutionary Strategies (ES) are a particular class of EAs, where the population is just one individual and only mutation operation is applied. Let  $\mathbf{X}$  be a chromosome defined by,

$$\mathbf{X} = (x_1, x_2, \dots, x_p; \sigma) \quad (4)$$

where  $x_i$  ( $i = 1, 2, \dots, p$ ) are the solutions parameters and  $p$  is the maximum number of parameters. Individuals could contain some strategy parameters of the mutator operator, for changed them during the run of the algorithm.

The mutation operation is defined applying a gaussian random number perturbation to chromosome parameters, generating a new mutated chromosome given by

$$\mathbf{X}' = (x'_1, x'_2, \dots, x'_p; \sigma'), \quad (5)$$

with

$$\sigma' = \sigma \cdot exp\left(\frac{1}{\sqrt{p}} \cdot N(0, 1)\right), \quad (6)$$

$$x'_i = x_i + N(0, \sigma'), \quad i = 1, 2, \dots, p, \quad (7)$$

where  $N(0, \sigma)$  is a gaussian distribution with zero mean and standard deviation  $\sigma$ . This  $\sigma$  is often called the mutation step size, it determines the extent to which given values  $x_i$  are perturbed by. By using a Gaussian distribution, small mutation are more likely than largest ones [3].

Theoretical studies motivated to use the 1/5 success rule of Rechenberg [3]. This rule is executed at periodic intervals

and states that the ratio of successful mutations to all mutations ( $f_s$ ) should be 1/5. Hence if the  $f_s$  is greater than 1/5 the  $\sigma$  should be increased to make a wider search of the space, and if the  $f_s$  is less than 1/5 then it should be decreased to concentrated the search more around the current solution, how described by equation 8.

$$\sigma = \begin{cases} \sigma/c & \text{if } f_s > 1/5, \\ \sigma \cdot c & \text{if } f_s < 1/5, \\ \sigma & \text{if } f_s = 1/5, \end{cases} \quad (8)$$

where the  $f_s$  is measured over a number os trials and the parameter  $c$  is the range  $0.817 \leq c \leq 1$  [3].

## 3. THE PROPOSED METHOD

The main characteristic of the GRASPES is based in the GRASP method, described on section 2.2, where it searches an ANN model for time series forecasting. The general expectative is that, given a sub-optimal solution, closed it there will, with high probability, other sub-optimal (or optimal) solutions. Consequently, the search will tend to look around of such solution, stopping when a local optimum model is found.

Based in the section 2.3, by definition, each individual codifies an three layer ANN MLP type. An ES initialize one individual  $I$ , which is a potential solution, usually generated randomly. The individual will be evaluated by a defined fitness function, described on section 4, where the better individuals will return higher fitness values. The population  $P_n$  is a set of  $I$ , where  $n$  is a user-defined parameter, such value is the quantity stored by the proposed method of the best individuals generated by the parent in all iterations.

The ES clones the parent's chromosomes  $I_p$  and will then undergo a operation of mutation which changes the genes of the chromosome, consequently, some features of the chromosomes inherited from their parent will be changed. For tuning the structure of the ANN, two integer random numbers  $l \in [1..10]$  are generated, one to define the ANN number of time lags (processing units in input layer  $i$ ) and another to define the number of processing units in hidden layer (sigmoidal units  $j$ ). For each weight of the optimal individual  $I$  a floating-point number  $e$  is randomly generated in the predefined interval  $T$ , with mutation step size  $\sigma$  (where the initial value is 1), governed by the law described by the equation 8.

The next step is add  $e$  to the weight as described on equation 7. This new individual is evaluated and will be included in the population  $P_n$  (equation 9), if and if only, its solution quality (measured by the fitness function as described on section 4) is better than the actual father. It is believed that high potential parents will produce better offspring (survival of the best ones).

$$P_n = [I_1, I_2, \dots, I_n], \quad n = pop\_size. \quad (9)$$

Then the process will continue with the procedure repetition: the parent's chromosomes is cloned and the operation of mutation is executed. This steps will be repeated until the mutated individuals number criterium or the size of the population  $n$  is reached. When this happens it will be said that a Parent's Generation (PG) occurs and the best individual of the current population  $P_n$  is selected, substituting the parent, as described on the equation 10.

$$I_p = max(f(P_n)) \quad (10)$$

The termination criterium is research when occurs a defined iteration number (PG) without better individual generation, where a new individual is considered "better" when your fitness is great than parent's fitness plus a constant  $u$ , i.e.,

$$Better\ Individual : f(Individual) > f(Parente) + u$$

## 4. PERFORMANCE EVALUATION

Five well known performance criteria should be considered for allowing a more robust performance evaluation, where  $T$  is the actual value of the time series data (target) and  $O$  is the model output (prediction):

*MSE (Quadratic Average Error):*

$$MSE = \frac{1}{N} \sum_{j=1}^N (e_j)^2, \quad (11)$$

where  $N$  is the amount of the target points on the set and  $e_j = (T_j - O_j)$ .

*MAPE (Percentage Average Error):*

$$MAPE = \frac{1}{N} \sum_{j=1}^N \left| \frac{e_j}{X_j} \right| \quad (12)$$

where  $X_j$  is the point on the set in the instant  $j$ .

*NMSE (Normalized Mean Square Error):*

$$Theil = \frac{\sum_{j=1}^N (T_j - O_j)^2}{\sum_{j=1}^N (T_j - T_{j+1})^2}. \quad (13)$$

*POCID (Prediction On Change Of Direction):*

$$POCID = 100 \frac{\sum_{j=1}^N D_j}{N}, \quad (14)$$

with

$$D_j = \begin{cases} 1 & \text{if } (T_j - T_{j-1})(O_j - O_{j-1}) > 0, \\ 0 & \text{other case.} \end{cases} \quad (15)$$

*ARV (Average Relative Variance):*

$$ARV = \frac{\sum_{j=1}^N (O_j - T_j)^2}{\sum_{j=1}^N (O_j - \bar{T})^2} \quad (16)$$

where  $\bar{T}$  is the average of the time series data.

For the best ANN model choice, it is calculated the fitness function of each individual based on the idea that the best results are those which have the smallest errors combined. Four fitness function were studied, which are:

$$fitness = \frac{POCID}{1 + MSE + MAPE + NMSE + ARV}, \quad (17)$$

$$fitness = \frac{1}{1 + MSE}, \quad (18)$$

$$fitness = \frac{POCID}{1 + MSE}, \quad (19)$$

and

$$fitness = \frac{POCID}{1 + NMSE}. \quad (20)$$

## 5. EXPERIMENTAL RESULTS

Two time series were used for evaluation of the GRASPES, a natural phenomena time series (Sunspot) and a financial time series (Dow Jones Industrial Average Index - DJIA). The investigated series were normalized to lie within the interval  $[0;1]$  and divided in three sets: training set with 50%

of the time series data, validation set with 25% of the time series data and test set with 25% of the time series data.

The MLP architecture is represented by  $i - j - 1$ , as described on section 2.1. The termination conditions for the algorithm as described on the section 3 are the  $u = 10^{-3}$  and  $PG = 150$ , the maximum size of the set  $P$  is  $n = 30000$  (it was observed that a parent don't generate more than 25000 offspring) and the maximum value of  $\sigma$  is 10.

In addition, experiments with the TAEF method (in-phase matching, using fitness function by the equation 18) found in [4] and MMNN found in [1] (using fitness function by the equation 17) are used for comparison with GRASPES.

### 5.1 Dow Jones Industrial Average (DJIA)

The Dow Jones Industrial Average (DJIA) Index series corresponds to daily records from January 1st 1998 to August 26th 2003, constituting a database of 1.420 points. For the prediction of the DJIA Index series the average of 10 experiments was done and the individual with the largest validation fitness function (Equation 17) is chosen to represent the model. The GRASPES automatically chose the architecture is 2-9-1 for the time series representation. Table 1 shows the results for all performance measures of the test set.

**Table 1: Results For The DJIA Index Series**

	TAEF	MMNN	GRASPES
ARV	0.0346	3.4423e-2	3.3300e-02
MAPE(%)	10.1529	9.67	9.6800
MSE	8.4183e-4	8.3236e-4	8.2345e-04
NMSE	1.0006	0.9945	0.9871
POCID	47.57	50.85	54.0845

### 5.2 Sunspot Series

The sunspot series used consisted of the total annual measures of the sun spots from the years 1700 to 1988, generating a database of 289 examples. N.Terui and H.K.Van Dijk [9] developed a work where a combination of some linear and non-linear models were employed for times series prediction. Among the series investigated, Terui and Van Dijk employed the sunspot series from the years 1720 to 1989 to test their method based on the combination of the AR, TAR and ExpAR models [9]. The best experimental results reported with their proposed method (best model combination) corresponded to an MSE error of 0.0390.

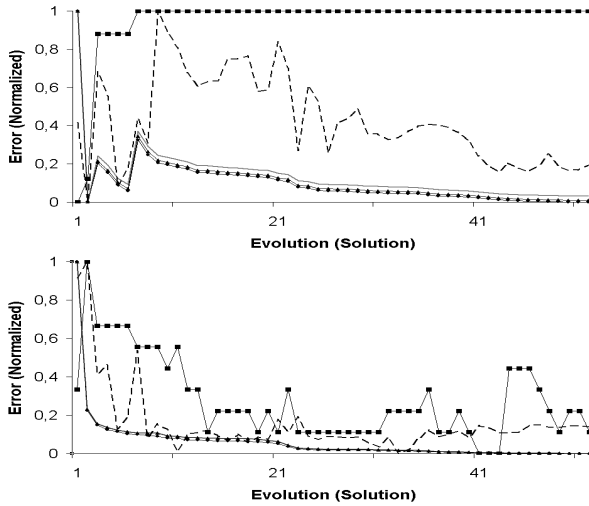
The experiments shows that some error functions are not necessarily interlinked, but others yes. Preliminary experiments were realized with the sunspot series. These experiments were realized with fitness functions described on section 4,  $n = 30000$ ,  $\sigma_{Max} = 10$ , five different initialization and showed that the interlinked errors are correlation with the dynamic of the optimal model search, i. e., as expected, for the same problem the choice of the fitness function affects the way that the offspring will evolve and, consequently, the final result. Most of the works found in the literature of time series prediction employ only MSE or NMSE error as performance criterion for model evaluation [1], but Table 2 with the results of the best individual found of the test set, show that it cannot be considered alone as a conclusive measure for comparison of different prediction models. It is important to note the correlation and variation of the error measures according with the fitness functions, i.e., small changes of the fitness evaluation can generate huge changes on the error values.

For the GRASPES driven by Equation 19, the Figure 5.2 shows an apparent behavior where the POCID and MSE evaluate measures have a direct influence about other calculated errors (MAPE, NMSE and ARV), while if the dynamic system is driven by fitness function Equation 20, as shows

**Table 2: Results For The Sunspot Series.**

	TAEF	GRASPES			
	Eq.18	Eq.17	Eq.18	Eq.19	Eq.20
ARV	0.123	0.135	0.128	0.843	0.097
MAPE(%)	30.06	29.88	37.85	136.25	33.86
MSE( $10^3$ )	7.000	8.775	8.179	5.545	6.300
Theil	0.176	0.329	0.269	2.079	0.238
POCID	84.058	72.222	50.684	97.142	73.611

in Figure 5.2, apparently, the MSE and POCID are mutually affected and the others calculated errors have a smooth evolution, independent of MSE and POCID's fluctuations.



**Figure 1: Behavior of Fitness Function-Equation 19 (superior) and 20 (inferior). Errors: ARV (solid line), MAPE (dashed lines), MSE (line with triangles), POCID (line with squares) and NMSE (line with circles).**

Another studies done in this work were the two types of approaching used, one leaves the coefficient  $\sigma$  co-evolving with the solutions freely without a maximum value, other uses a maximum value to narrow the area of search with  $\sigma_{Max} = 10$ . All the best values showed in the tables 1 and 2 has founded with a maximum  $\sigma$ . The average of the experimental data are changed when the  $\sigma$  has a maximum value, being the fitness average greater when  $\sigma$  is free. In order to show that the cases are statistically different, not only different about the fitness average value, it will be used the test of zero mean described by Jain [5]. Using the obtained result on the DJIA experiments with confidence of 95%, the confidence interval calculated is  $(-3.8145; -0.9971)$ , which not includes zero. Therefore, by the cited test, in this case the two systems are different.

With this results another doubt appears, how this maximum value of  $\sigma$  could affect the finals results? In order to clarify this point other tests were made, this time with the  $\sigma$  with a set of maximum value  $\{1, 5, 10, 50, 100, 500\}$ . Five experiments was realized with  $n = 30000$  and each value of  $\sigma$  maximum. According with the obtained results the average of fitness value decrease with the increasing of the maximum  $\sigma$ . Applying again the test of zero mean with confidence of 95%, the confidence interval calculated is  $(1, 2282; 12, 4452)$ , which not includes zero, i.e., the two systems are different, and consequently, is better to search the state space area with no longer step sizes.

Finally, the variation of results according with the size  $n$  of the parent set  $P_n$  seems change. Ten experiments were made with range  $n=\{1, 10, 100, 1000, 10000, 30000\}$  and it

was observed that the values of the mean of the samples has a tendency of increase when increases the  $n$  too, which is not so obvious, because when  $n$  increases, the number of chances to get sub-optimal values increases too [3]. Applying the test of zero mean to check the differences between two different sets of  $P_n$ , one with  $n = 100$  and another with  $n = 30000$  and the  $\sigma$  free, when the confidence is 95%, the confidence interval calculated is  $(-4, 4646; 0, 8488)$ , which includes zero. In this case, the two systems are not different.

## 6. CONCLUSIONS

This paper has presented a study for multi-objective fitness function and evolution parameters for time series forecasting with an intelligent hybrid system called GRASPES.

The experimental results using five different metrics (MSE, MAPE, NMSE, POCID, ARV) showed that small changes of the parameters can boost the performance of time series prediction. The experimental validation of the method was carried out on some complex and relevant time series and were compared to other methods like TAEF Method [4] and MMNN [1] with the same time series and shown how delicate and important are the changes of the fitness evaluation.

Future works will consider: test with other time series; the phase prediction adjust procedure [4]; the use of MLPs convolutional training algorithms; the other distinct forms of modeling the ANN for this problem; a deeper study evolving the behavior of the  $\sigma$ ; and the different fitness functions.

## 7. ACKNOWLEDGMENTS

I would like make a special thanks to my advise Professor Tiago A. E. Ferreira (Statistic and Informatics Department of the Rural Federal University of Pernambuco) by believe in my potential and that scientific task is made with hard work and not with grades.

## 8. REFERENCES

- [1] R. Araujo, F. Madeiro, R. de Sousa, L. Pessoa, and T. Ferreira. An evolutionary morphological approach for financial time series forecasting. *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pages 2467–2474, 16-21 July 2006.
- [2] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematical Control Signals Systems*, 2:303–314, 1989.
- [3] A. Eiben and J. Smith. *Introduction to Evolutionary Computing*. Springer, 2003.
- [4] T. Ferreira, G. Vasconcelos, and P. Adeodato. A new evolutionary method for time series forecasting. In *GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation*, pages 2221–2222, New York, NY, USA, 2005. ACM.
- [5] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, INC, New York, 1991.
- [6] J. Koza. *Genetic Programming - On the Programming of Computers by Means of Natural Selection*. The MIT Press, Cambridge, Massachusetts, 1998.
- [7] M. Resende and C. Ribeiro. Greedy randomized adaptive search procedures. In *Handbook of Metaheuristics*, volume 57, pages 219–250. Springer, 2003.
- [8] F. Takens. Detecting strange attractor in turbulence. In A. Dold and B. Eckmann, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381, New York, 1980. Springer-Verlag.
- [9] N. Teru and H. Dijk. Combined forecasts form linear and nonlinear time series models. *International Journal of Forecasting*, (18):421U438, 2002.