X Congresso Brasileiro de Inteligência Computacional (CBIC'2011), 8 a 11 de Novembro de 2011, Fortaleza, Ceará

TESTS WITH DIFFERENT FITNESS FUNCTIONS FOR TUNING OF ARTIFICIAL NEURAL NETWORKS WITH GENETIC ALGORITHMS

Aranildo R. Lima[a], Paulo S. G. de Mattos Neto[b]
[a]Department of Earth and Ocean Sciences – University Of British Columbia - Vancouver - Canada
[b]Center for Informatics – Federal University of Pernambuco - Recife - Brazil
proj.brain@gmail.com, psgmn@cin.ufpe.br

David A. Silva and Tiago A. E. Ferreira
Statistics Department and Informatics Department – Federal Rural University of Pernambuco - Recife - Brazil
expressosub@yahoo.com.br, tiago@deinfo.ufrpe.br

Abstract – The choice of a good fitness function still a key element for the practitioners who use artificial intelligence to solve the forecasting problem. The fitness functions proposed in the literature have not been compared among them. Based on this fact, we started a brief empirical comparison among three different fitness functions in order to give some guidelines to help the fitness function choice. They were tested with a modified Genetic Algorithm for tuning the Artificial Neural Network structure and parameters. This experimental investigation with six non linear time series, showed that adjust the fitness function can be lead to a significantly improved accuracy for one given performance measure.

Keywords – Time Series Forecasting, Genetic Algorithms, Fitness Function, Artificial Neural Networks and Artificial Intelligence.

1 Introduction

Some of the most promising approaches for forecasting are based on Artificial Intelligence context. For instance, Artificial Neural Networks (ANNs) have been successful applied for the nonlinear modeling of time series [1]. However, for the time series forecasting problem, ANNs use a set of adjustable parameters including: network topology, number of processing units, etc. In many instances determine the optimal or sub-optimal values of these parameters is a difficult task. To set up all these parameters some hybrid intelligent techniques based on Evolutionary Algorithms (EA) have been used [2–8]

In EA the process of biological evolution is mimicked and the natural selection concept is applied to a pictorial population (each individual of the population is a feasible problem solution). Based on a Fitness Function (*FF*), the fittest individuals are chosen to seed the next generation of population, evolving the population to the optimal solution. In this way, designing the evaluation function is crucial because *FFs* often are the only information about the problem in the algorithm [9],usable knowledge about the problem domain should be used [9].

Taking a brief look into the previous studies is easy to see the heterogeneity of *FFs* for the same problem and goal (better accuracy by decreasing the error measures). Most of these studies use the conventional *FF* with Mean Squared Error (MSE) [10], but also can be found with Root Mean Square Error(RMSE) [7], Mean Absolute Percentage Error(MAPE) [11], Normalized Root Mean Square Error (NRMSE) [3] (which in this particular case is described by the authors as "the obvious choice"), Average Relative Variance (ARV) [8] also called Normalized Mean Square Error(NMSE) and complexes functions mixing some of these errors [6].

It seems that the choice of a good or "obvious" *FF* still an open question for the practitioners who use EA to solve the forecasting problem. The purpose of this work is to introduce and provide some guidelines about *FF* and Hybrid Methods for the time series forecasting. Comparing empirically the commonly adopted (and introducing a new one, Equation 5) *FFs* and using a modified Genetic Algorithm (GA) [4,5] we begin to indicate some important aspects (effectiveness and efficiency) such as avoid local minimal, overall predictions and best accuracy.

Section 2 has a brief explanation about the Time Series Forecasting Problem and Artificial Neural Networks. Section 3 provide the main steps of the modified GA. Section 4 gives a description of statistical error measures and the tested *FFs*. Sections 5 and 6 show the experimental results and conclusions respectively.

2 Time Series Problem and ANN

In the branch of statistics, signal processing, or many other study fields, a time series is a set of data points S_t, measured generally at successive times, spaced at (often uniform) time intervals defined by,

$$S_t = \{s_t \in \mathbb{R} \mid t = 1, 2, 3 \ldots N\}, \quad (1)$$

where t is the temporal index and N is the number of observations.

The objective of the time series forecasting techniques is to identify patterns presents the data and to build a model able to identify the next time patterns. Often a non-trivial problem, considering that some time series can have many types of components, such as trends, seasonality, impulses, steps, model exchange and other non-linearities.

The ANNs have powerful pattern classification and pattern recognition capabilities. For this reason one major application area is forecasting. They provide an attractive alternative tool for both forecasting researchers and practitioners [1].

As proven by the Cybenko theorem [12], a MultiLayer Perceptron (MLP) ANN with at least one hidden layer is capable of approximating any continuous function. The purpose of this work is predict continuous functions, then MLP networks with one hidden layer and architecture I-J-K were used. The I denotes the number of time lags (processing units in input layer), J denotes the number of processing units in hidden layer with sigmoidal activation function (Sig) [13] and K denotes the number of processing units in output layer.

The output of ANN is given by,

$$y_k(t) = \sum_{j=1}^{J} W_{jk} Sig\left[\sum_{i=1}^{I} (W_{ij} Z_i(t) - b_j^1)\right] - Sig(b_k^2), \tag{2}$$

where $Z_i(t)$ ($i = 1, 2, \dots, I$) are the ANN input values. Since the prediction horizon is one step ahead, only one output unit is necessary ($k = 1$).

The other parameters of Equation 2 are:

- W_{ij}, weights of connections of the input layer to the hidden layer;
- W_{jk}, weights of connections of the hidden layer to the output layer;
- b_j^1, bias of the hidden unit;
- b_k^2, bias of the output unit,

and all these parameters are real values.

3 The Modified GA and FF

Improving the ANN prediction performance can be achieved through the correct adjustment of its parameters. In other hand, the ANNs parameters are problem dependent and the procedure to adjust them demand a search into a very wide space.

Genetic Algorithms (GA) [14] are a well-known technique of directed random search widely applied in complex optimization problems. They are particularly attractive to use in situations where the number of parameters is very large and analytical solutions are very difficult, or impossible, to obtain. The modified GA used in these experiments was originally proposed by [4] where new genetic operations were introduced to improve its performance [5].

In GAs the population is composed by set of trial solutions of the problem. Each solution (individual) is coded by an appropriate data structure (often a parameter vector) referred to as chromosome and evaluated by a *FF*.

Let **X** be a chromosome defined by,

$$\mathbf{X} = (x_1, x_2, \dots, x_m) \tag{3}$$

where x_i is a parameter of solution, with $i = 1, 2, \dots, m$, and m is the maximum number of parameters. Equation 3 represents the chromosome used to describe a three-layer ANN parameters, coded as $[W_{ij},\ W_{jk},\ b_j^1,\ b_k^2]$ (Section 2).

In general, a problem solver is a step-by-step strategy, or method, intended for problem solving. Direct and indirect methods are commonly distinguished. A direct method is deductive, i.e. it starts directly from the input data and works its way towards the true solution through mathematical-analytical steps. However, when a problem is "ill-behaved" (which might be the case when it is nonlinear or discontinuous) such that direct approaches are not feasible, one will need to resort to indirect methods. Indirect methods are iterative procedures that make a sequence of supposedly improving guesses at the true solution in the problem space by guided trial and error. This can only be accomplished by way of an optimality criterion or evaluation criterion.

We first assume a non-negative-real valued scalar *FF* $f(s)$ overall solutions $s \in S$, which is a ground set $E = e_1, \dots, e_n$, a set of feasible solutions $S \subseteq 2^E$ and $f : 2^E \to R_+^*$. The ground set E, the *FF* f, and the constraints defining the set of feasible solutions (also called the search space) S are defined and specific for each problem. We seek an optimal solution $s^* \in S$ such that $f(s^*) \leq f(s), \forall s \in S$.

The chain of guesses made at the true solution during optimization may be construed as a search trajectory - the path that is traversed through the search space (different *FF* means different path). Basically, the *FF*, assigns a fitness value to each point in the parameter space [15], where this value can be seen as a measure of how good a solution, represented by that point in the landscape, is to given problem [16]. Accordingly, the combination of the search space and the *FF* results in an observation landscape. Assuming the goal is to maximize fitness, we can imagine the globally best solutions (the global optima, or "globals") as "peaks" in the search space. Optimization thus comes down to hill-climbing on the observation landscape, with the intent of finding the highest peak. All optima that are not a global optimum are called sub-optima or local optima, and the candidate solution that corresponds with a local optimum is called a local solution [17], available and usable knowledge about the problem domain should be used [9]. Therefore, the correct choice of the *FF* is crucial for a good solution of the problem.

In doing so the trials *FFs* are:

$$f_1(I) = \frac{1}{1 + MSE}, \tag{4}$$

$$f_2(I) = \frac{1}{1 + THEIL}, \tag{5}$$

$$f_3(I) = \frac{1}{1 + ARV}, \tag{6}$$

where the error measures are describe in Section 4.

The GA stopping criterion are: training progress (the GA stops when occurs a defined number of generations without a perceptual increase the average of the population quality); the maximum number of the GA generations; and, loss generation (a decline of mean quality of the population about the validation set). The main steps necessary for implementing the modified GA are in Algorithm below, Figure 1.

Figura 1: Procedure of the Modified GA

begin
- $\tau \to 0$; `// τ: Number of iteration;`
- initialize $\mathbf{Pop}(\tau)$; `// Pop(τ):population for iteration τ;`
- evaluate $f(\mathbf{Pop}(\tau))$; `// f(Pop(τ)): fitness function;`
- **while** *not termination condition* **do**
  - $\tau \to \tau + 1$; ;
  - Select two parents $\mathbf{p_1}$ and $\mathbf{p_2}$ from $\mathbf{Pop}(\tau)$;
  - Perform *crossover* operation ;
  - Perform *mutation* operation ;
  - `// Reproduce a new Population;`
  - The chromosome generated by the crossover and mutation operation with the largest fitness value replaces the chromosome with the smallest fitness value in the $\mathbf{Pop}(\tau - 1)$;
  - evaluate $f(\mathbf{Pop}(\tau))$; ;
- **end**

end

More information about this modified GA and its implementation for the time series forecasting problem with the ANNs of MultiLayer Perceptron (MLP) types and its used parameters are provided in [4, 5, 10].

4 Performance Evaluation

In evaluation of forecasting accuracy, including forecasting competitions, researchers have paid attention to the selection of time series and to the appropriateness of forecast-error measures [18]. The use of only one error for evaluate the model performance does not shows the performance of the prediction in a clear way [19]. None of the error measures is the best on all criteria (as cost, reliability, sensitivity to small changes, protection against outliers, relationship to decision making, etc) [20]. Therefore, which is the best error measure to characterize a given problem? The selection of an error measure is dependent upon the situation. For example, the turning point is the most important one when the prediction is used for judging the economical time series to sell or buy. In all mentioned works (Section 1) and in this paper, the main goal is best accuracy of the model toward the error used in the *FF*.

In this way, four well-known error measures are considered to a robust evaluation of the prediction performance.

***MSE** (Mean Squared Error)*:

$$MSE = \frac{1}{N} \sum_{j=1}^{N} \left(T_j - O_j\right)^2, \tag{7}$$

***THEIL**(U of Theil Statistics)*:

$$Theil = \frac{\sum_{j=1}^{N} \left(T_j - O_j\right)^2}{\sum_{j=1}^{N} \left(T_j - T_{j+1}\right)^2}. \tag{8}$$

***ARV** (Average Relative Variance)*:

$$ARV = \frac{\sum_{j=1}^{N} (O_j - T_j)^2}{\sum_{j=1}^{N} (O_j - \overline{T})^2} \tag{9}$$

where T is the value to be predicted (target), O is the model output (prediction), N is the amount of the target points and $\overline{T}$ is the average of time series. In an ideal model all the performance measures will tend to zero.

As the objective of this work is to give an overall idea of the *FFs* performance and accuracy, we will assume for benchmark only the MSE which is the most popular measure used for performance prediction. Much has been written about the choice of forecast-error statistics. A good overview is provided in a series of articles and commentaries in the International Journal of Forecasting [18–20].

4.1 Series Used

Conclusions about the accuracy of various forecasting methods typically require comparisons across some time series [20]. Six complex time series were used as *FFs* benchmark: an artificial, Henon Map [21, 22]; two natural phenomena, Sunspot and Star Series (data from *http://robjhyndman.com/TSDL/*); three financial, Standard & Poor 500 (S&P500) Stock Index, Petrobras Stock Values and Dow Jones Industrial Average Index (data from *http://finance.yahoo.com*).

The Henon series (HENON) is a relevant artificial time series due to its complex nature and chaotic dynamics. The series constitute a database of 1000 points and is given by

$$X_t = 1 - a(X_{t-2} - r_{t-2})^2 + b(X_{t-4} - r_{t-4} + r_t) \tag{10}$$

where $a = 1.4$, $b = 0.3$ and $r_t = 0$ (generated without the inclusion of any noise) [22].

The sunspot (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. Solar activity is difficult to predict using standard models due to high frequency content, noise contamination, high dispersion level, etc [8].

The Star series (STAR) corresponds to a magnitude of an oscillating shine star, observed daily in the same place and hour, constituting a database of 600 points.

The S&P500 Stock Index is a pondered index of market values of the most negotiated actions in the New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and Nasdaq National Market System. The S&P500 series corresponds to the monthly records from January 1970 to August 2003, constituting a database of 369 points. To reduce exponential trend the natural logarithm was applied to the original values of this series.

The Petrobras Stock Values series (PETRO) corresponds to the daily records of Brazilian Petroleum Company from January 1st 1995 to July 3rd 2003, constituting a database of 2,060 points.

The Dow Jones Industrial Average Index (DJIA) series corresponds to daily observations from January 1st 1998 to August 26th 2003, constituting a database of 1,420 points.

All series investigated were normalized to lie within the interval [0;1] and divided in three sets: training set (50% of the data), validation set (25% of the data) and test set (last 25% of data). For the purposes of this paper which the focus is the *FFs*, the window lag for the time series representation and the J-layer size used were previously studied by Ferreira [10, 23] (I-J-K as described on Section 2).

The GA (described in the Section 3) parameters used in this work were based on previous research in literature [4, 5]. A sample is started with a population composed by 20 individuals initialized randomly and stops when any of the stopping criterion are reached. The GA stopping criterion were: 5000 generation; 1500 continuous generation without increasing of 1% the mean quality of population; and decreasing of 10% toward the best value of population average quality (about the validation set) already achieved. The Best Individual (or just BI for brevity) of the population according to the best value (of test set) of the studied *FFs* was chosen to represent the model created by the sample. For each time series with a specific *FF*, thirty samples were repeated. Then each MSE value of the BIs were used to represent the distribution of probability and to construct the Boxplot. Boxes extend from the 25th to 75th percentiles, with the line indicating the median. Whiskers represent the most extreme data within ± 1.5 times the interquartile range (i.e. the box height); values outside this range are plotted as dots (outliers)

5 Experimental Results

Figure 2 shows the MSE results of the DJIA, PETRO and S&P500 series. For the boxplot the considerations are:

1. If the confidence intervals do not overlap. The alternative with higher sample mean is significantly worse.

2. If the confidence intervals overlap considerably such that the mean of one falls in the interval for the other. The two alternatives are equal with the desired confidence.

3. If the confidence intervals overlap slightly such that the the mean of either is outside the confidence interval for the other. In this case, no visual conclusion can be drawn. It is necessary to do another test. Another way is looking the notches. If the notches of two plots do not overlap this is strong evidence that the two medians differ.

In this way, all the alternatives can be considered equals even if f_1, visually, seems sightly better than f_2 and f_3 in the DJIA case. However, f_2 has less dispersion (for the forecasting problem more dispersion is considered good if is skewed to zero) than the others *FFs* emerging as a good option of *FF* to be use in this time series, even with its two outliers which one of them lies within the confidence interval of the f_1 and f_3.

The main point of the Figure 2 is the fact that the use of the error measure inside of *FF* does not guarantee the best results of that error, *i. e.*, just put the MSE inside the *FF* will not give the best MSE results at the end. This fact is happening again with PETRO series. Despite the notch of the three boxplots are almost at the same region, the dispersion and confidence intervals of f_2 and f_3 are not just smaller, but also closer to the zero which indicates better predictions. The same behavior can be found at S&P500 series but this time just subtly.

For Figure 3 the boxplot of f_3 is not showed because its values were too high in comparison with the others *FFs* and add its boxplot turns the comparison between f_1 and f_2 confused. The results of STAR series has similar characteristics than the previous

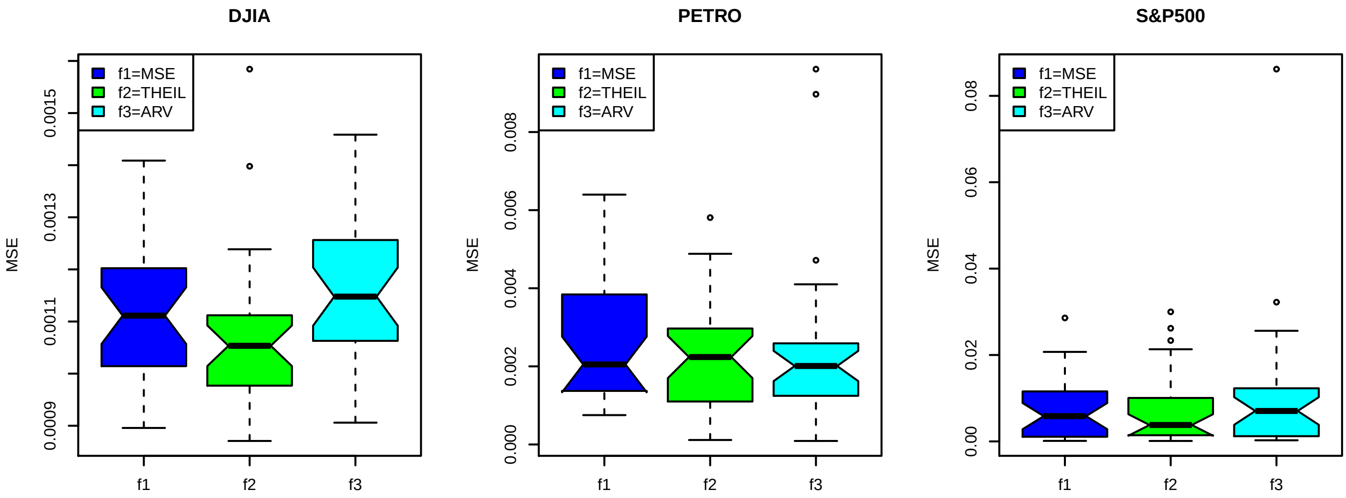

Figura 2: Boxplot of the MSE results for the DJIA, PETRO and S&P500 series using f_1, f_2 and f_3 as fitness functions.

series. However the interesting point now is not only the less variability using f_2 but also the number of results considered bad outliers (bad because they lies outside the upper confidence region which for the forecasting problem is considered a bad result) at the f_1 boxplot.

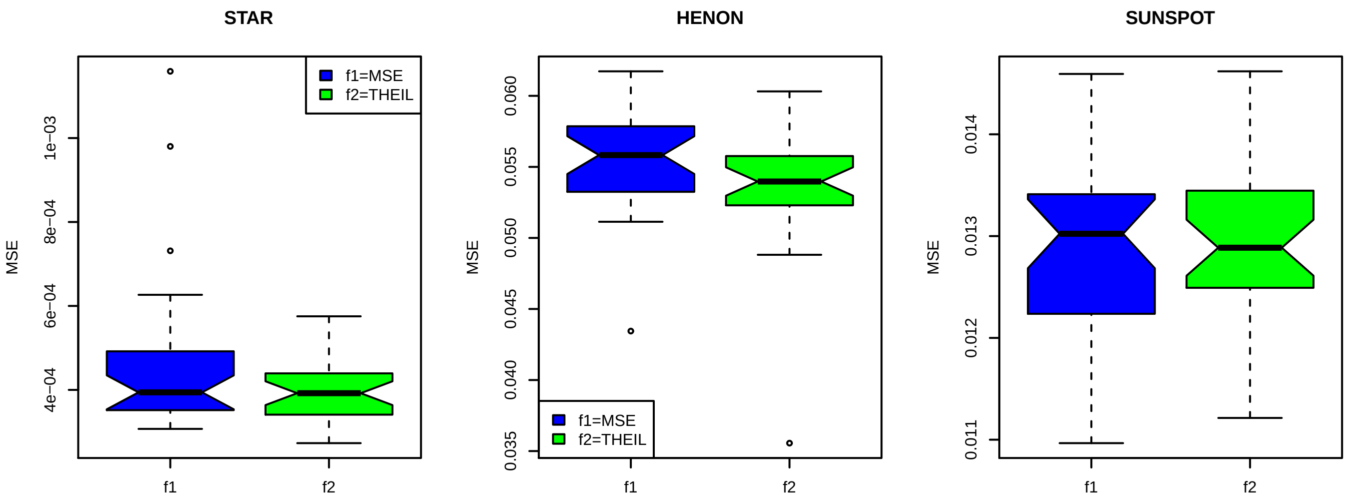

Figura 3: Boxplot of the MSE results for the STAR, HENON and SUNSPOT series using f_1 and f_2 as fitness functions.

The boxplot related to HENON series show to us in a more clear way the report made before about the error measure inside the *FFs*. Despite the fact that again f_2 has less dispersion related to the top (bad dispersion), both functions have outliers. Those outliers represent the best results of all the samples. It is easy to see that f_2 using THEIL error as objective achieved the best MSE result than f_1 which uses MSE error as objective. As commented at Section 3, each *FF* one possible explanation about what is really happening is relate to the ability of the *FF* get stuck or not at the local optimum. Another one is the constructed trajectory through the search space, *i. e.*, it is possible that the path used by f_2 has less chances to get stuck at local optimum than the path used by f_1. The path and local optimum can be sometimes highly dependent of the used time series. For example, with SUNSPOT series f_1 has better dispersion (tends to the bottom) than f_2 and this is the only case when it is happening.

Talking about f_3 it seems that its ability to get stuck inside a local optimum (or path) has more sensitivity to the time series used than the others *FFs*. The series showed in Figure 2 are financial which have different characteristics than the series showed in Figure 3. Those characteristics have more impact in the f_3 results than the others *FFs* and they are statistically relevant.

A important question can be asked about the convergence. Among the three functions, which is the best in therms of convergence? How many iterations are necessary to reach the stopping criterion of the algorithm? Note that the same conditions are applied to all the *FF*s. Figure 4 shows the boxplot of last good generation recorded, *i. e.*, the generation number which has the last improvement in therms of *FF* average (related to each sample). Now it is clear that f_1 can't be beat. Its convergence is faster than the others *FFs*. Sometimes f_2 uses the double of the iterations.

However a trick observation can be made. According to the GA stopping criterion the maximum is 5000 generations. If the last generation recorded was 3501 for example, the algorithm will stop before the stopping criterion of 1500 generations without improvement. Because this case occurs very often a question can arise: Was f_2 really able to reach the optimum or if we will relax the stopping criterion of maximum iterations this function will be able to achieve better results? This definitely should be considered in future works.

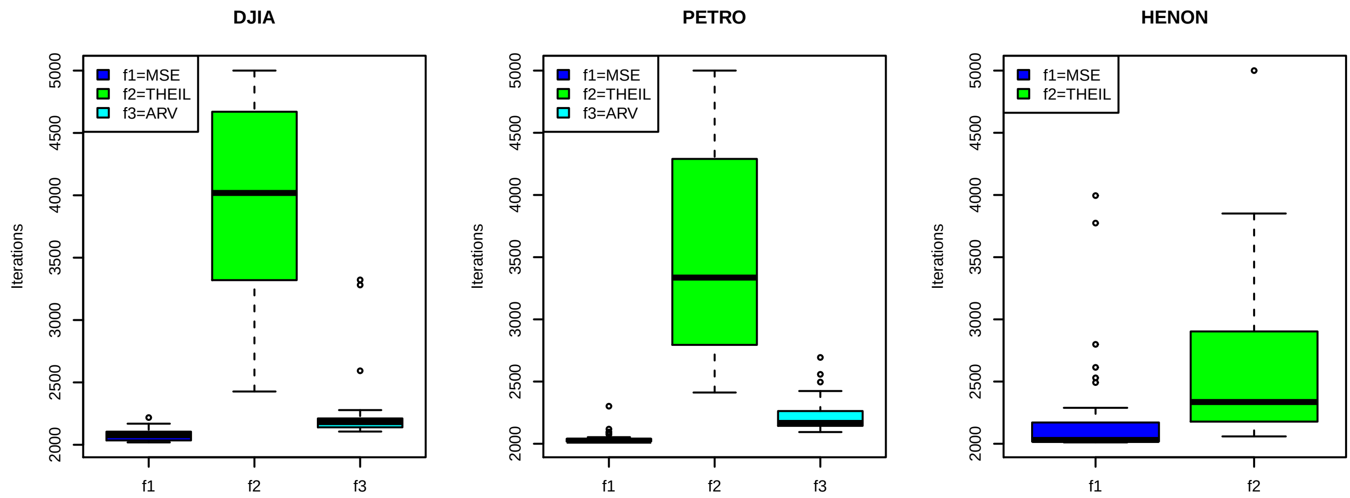

Figura 4: Boxplot of the amount of iterations MSE results for the STAR, HENON and SUNSPOT series using f_1 and f_2 as fitness functions.

In order to illustrate how can be the forecasting with different *FFs* we used the BI according with the highest value of its characteristic *FFs*. The figures 7, 5 and 6 show the forecasting of the S&P500 series. Clearly f_1 and f_2 have better prediction than f_3. Looking carefully the prediction, mainly between the range of 60 and 100 (Figures 5 and 6), an adequate choice to reach precise predictive model is f_2.

6 Conclusions

This paper has presented a study of *FFs* for time series forecasting using a GA to train a population of MLP networks.

The experimental results using three different metrics (MSE, THEIL and ARV) showed that changes of the fitness parameters for a same method can boost the performance of time series prediction. It can be conclude that it is possible to adjust only the *FF* (and must be careful selected), based on the error measure, to reach better predictions results. The experiments show that the use of THEIL error in the *FF* should be consider in future works of time series forecasting due its accuracy.

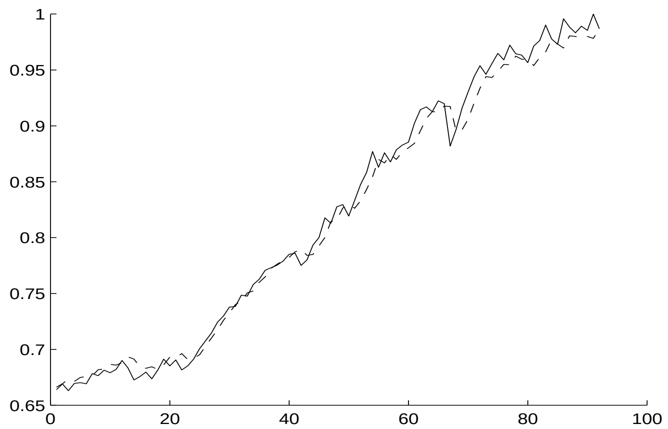

Figura 5: Results of BI with highest *FF* according f_1 for the S&P500 series - Axis:Y - Normalized Index, X - Monthly Records

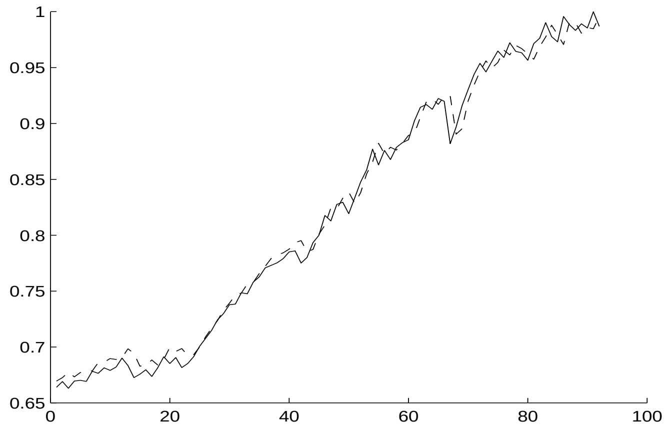

Figura 6: Results of BI with highest *FF* according f_2 for the S&P500 series - Axis:Y - Normalized Index, X - Monthly Records

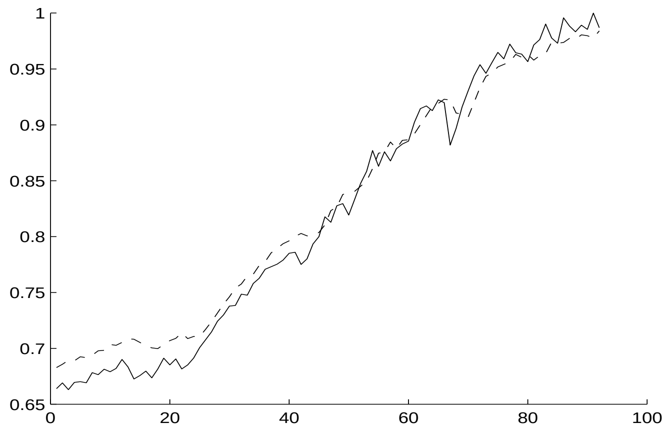

Figura 7: Results of BI with highest *FF* according f_3 for the S&P500 series - Axis:Y - Normalized Index, X - Monthly Records

The experiments show that using determined error measure at the *FF* will not guarantee the best results of this error measure. In fact the obtained results suggest that statistical errors measures in *FF* have a non-linear interrelationship and as expected the choice of *FF* is not a trivial decision.

Again is important to remember that each forecasting always depends of the goals. The results presented here give a guideline to a desirable choice of *FF* for a determined purpose. This purpose could be for example turning point, measure trending, overall performance, best accuracy, fastest convergence etc. In this study the goal is to minimize the error measures values in general. The uses of f_1 is recommended for a rapid convergence with a good overall performance. The f_3 function has a lot of variability related to the time series, so maybe is not a good option. When the final goal is not the sub-optimal adjustment, but the optimal adjustment (better accuracy), therefore the more appropriate *FF* is f_2. Despite the fact that the median are considered statistically equals the dispersions are less and closest to zero when compared with f_1. In other hand f_2 is the worse function in therms of convergence.

In this work was used three previously studied time series. However, a further study is being conducted to determine the possible limitations of the error measures and *FFs* utilized when dealing with other types of components also found in real world time series, such as trends, seasonality, impulses, steps, model exchange and other non-linearities.

Meanwhile, another study is being conducted with others hybrid system which uses Evolutionary Algorithms and the preliminary results shows that *FF* choice have more dependency for the problem (in this particular case, the time series) than the applied method in quantitative sense, but in qualitative sense the general aspects are very similar as found here. To aboard these time series features, a new approach of Artificial Intelligence is in development, to create an evolutionary *FF* (co-evolution process), where error measures could be dynamically combined in the same *FF*.

Future works will consider: other error measures [20]; other types of time series analysis [18]; the use of MLPs modified convectional training algorithms [13]; combinations with other Evolutionary Strategies, Genetic Programming [9] and hybrid methods (as PSO [6]); and extend this approach with others problems which uses evolutionary computing as fingerprint and image recognition.

REFERENCES

[1] 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.

[2] A. Rodrigues, T. Ferreira and R. de A. Araujo. "An experimental study with a Hybrid method for tuning neural network for time series prediction". pp. 3435–3442, June 2008.

[3] V. M. Landassuri-Moreno and J. A. Bullinaria. "Neural network ensembles for time series forecasting". In *GECCO '09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pp. 1235–1242, New York, NY, USA, 2009. ACM.

[4] F. Leung, H. Lam, S. Ling and P. Tam. "Tuning of the structure and parameters of a neural network using an improved genetic algorithm". *Neural Networks, IEEE Transactions on*, vol. 12, no. 1, pp. 79–88, Jan 2003.

[5] A. Rodrigues, P. S. G. de Mattos Neto and T. A. E. Ferreira. "A Prime Step in the Time Series Forecasting with Hybrid Methods: The Fitness Function Choise". In *Proceedings of International Joint Conference on Neural Networks*, Atlanta - Georgia, 2009. IEEE.

[6] P. S. G. de Mattos Neto, G. G. Petry, A. Rodrigues and T. A. E. Ferreira. "Combining Artificial Neural Network and Particle Swarm System for Time Series Forecasting." In *Proceedings of International Joint Conference on Neural Networks*, Atlanta - Georgia, 2009. IEEE.

[7] "Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds". *Neurocomputing*, vol. 73, no. 1-3, pp. 49 – 59, 2009.

[8] S. Marra and F. Morabito. "Solar Activity Forecasting by Incorporating Prior Knowledge from Nonlinear Dynamics into Neural Networks". pp. 3722–3728, 0-0 2006.

[9] A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Natural Computing Series. Springer, Berlin, 2003.

[10] T. A. E. Ferreira, G. C. Vasconcelos and P. J. L. Adeodato. "A New Intelligent System Methodology for Times Series Forecasting with Artificial Neural Networks". *Neural Process Letters*, vol. 28, no. 2, pp. 113–129, 2008.

[11] "Evolving artificial neural networks for short term load forecasting". *Neurocomputing*, vol. 23, no. 1-3, pp. 265 – 276, 1998.

[12] G. Cybenko. "Approximation by superpositions of a sigmoidal function". *Mathematical Control Signals Systems*, vol. 2, pp. 303–314, 1989.

[13] S. Haykin. *Neural Networks - A Comprehensive Foundation*. Pearson Education, second edition, 1998.

[14] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

[15] A. E. Eiben and M. Schoenauer. "Evolutionary computing". *Information Processing Letters*, vol. 82, no. 1, pp. 1–6, 2002.

[16] W. Hordijk. "A Measure of Landscapes". *Evolutionary Computation*, vol. 4, no. 4, pp. 335–360, 1996.

[17] "Understanding and using genetic algorithms Part 1. Concepts, properties and context". *Chemometrics and Intelligent Laboratory Systems*, vol. 19, no. 1, pp. 1 – 33, 1993.

[18] L. J. Tashman. "Out-of-sample tests of forecasting accuracy: An analysis and review". *International Journal of Forecasting*, vol. 16, no. 4, pp. 437–450, 2000.

[19] M. P. Clements and D. F. Hendry. "On the limitations of comparing mean square forecast errors". *Journal of Forecasting*, vol. 12, no. 8, pp. 617–637, Dec 1993.

[20] J. S. Armstrong and F. Collopy. "Error measures for generalizing about forecasting methods: Empirical comparisons". *International Journal of Forecasting*, vol. 8, no. 1, pp. 69–80, 1992.

[21] M. Henon. "A two-dimensional mapping with a stranger attractor". *Communications in Mathematical Physics*, vol. 50, no. 1, pp. 69–77, 1976.

[22] D. B. Murray. "Forecasting a chaotic time series using an improved metric for embedding space". *Physica D*, , no. 68, pp. 318325, 1993.

[23] T. A. E. Ferreira. "A new Intelligent Hybrid Methodology for the Time Serie Forecasting". Ph.D. thesis, Federal University of Pernambuco, 50732-970, Recife-PE, Brazil, Febuary 2006.