An Efficient Approach for Solving the 11 and the 111 Time Series Forecasting Competition Problem

Paulo J. L. Adeodato, Germano C. Vasconcelos, Adrian L. Arnaud, Rodrigo C. L. V. Cunha, Domingos S. M. P. Monteiro

***Abstract*— This paper presents a solution for the NN3-Forecasting Competition for both the 11 and 111 time series problems. Initially, the last 18 data examples from each series were separated for producing a completely independent test set. The remainder of each series was then filtered for trend elimination with the best fitting linear function. Correlation and Fourier analysis were used for the identification of the relevant time lags for each individual series. Careful normalization and temporal preprocessing were conducted on all the data used for modeling. Multilayer perceptron networks (MLP) were applied as the predictive technique due to its quality and robustness. Optimization was carried out on the MLP architecture as well as on its training algorithm through an optimization procedure focused on the best median of the SMAPE metrics on the validation set and on the minimum difference for the same metrics on the test data. An output phase correction scheme was employed to adjust any possible time phase distortions that appear in some series. The choices for network configuration ranged from 1 to 30 hidden neurons and the training algorithm was either the standard error backpropagation or Levenberg-Marquadt. The data for the 18 steps ahead were then produced by 15 networks for each time series and their median value was chosen at each time. Afterwards, the values of each predicted series were de-normalized by the appropriate inverse factor and had its cubic polynomial trend re-inserted for producing the 18-month-ahead forecast. Experiments showed that the solutions worked well for most of the series each having its own architecture, algorithm and output phase correction.**

I. Introduction

DESPITE all the investment seen on research for better neural network solutions to time series prediction, their performance is very far from those already achieved on other non-deterministic problems (e.g. classification, optimization etc.). Also, neural network solutions still lag behind other traditional approaches when dealing with multiple time series forecasting [1]. In this context, several attempts have been made with different artificial neural network paradigms as part of evolutionary or hybrid systems optimized for predictive applications.

Several different approaches have been proposed for time series prediction. The statistical technique of Box & Jenkis (ARIMA models) [2] became one of the most popular among practitioners in actual world forecasting tasks. However, ARIMA models are linear, a feature which represents a limitation for predictive modeling. Nonlinear approaches have been proposed for overcoming this constraint. Bilinear models [3], threshold autoregressive models [4] and exponential autoregressive models [5] among others are examples of such attempts. These nonlinear approaches, however, are mathematically very complex. Artificial neural networks were a recent alternative proposed for non-linear modeling of time series [6], more recently combined with evolutionary approaches for the network parameters' optimization (e.g. topology, number of processing units, learning rate etc.) [7].

The strength of the work carried out here, however, relies strongly on an ensemble of ideas from different areas, more on its sound statistical procedures, and less on the particular forecasting techniques used.

The summary of the steps carried out in each series is listed below. The idea is focused on a systematic approach for predicting multiple time series:

1. Test set separation
2. Series analyses
3. Trend removal
4. Data normalization
5. Predictive model optimized selection
6. 15-Replicas model training without test set
7. Median model selection
8. Performance evaluation
9. 15-Replicas model training with test set
10. Median forecast selection
11. Data de-normalization
12. Trend re-insertion

This paper is organized as follows. Section II presents the data selection approach. In Section III, the data preparation and analysis are explained. Section IV describes the predictive modeling. Section V presents some results and interpretation on the test data. Section VI finalizes the paper with remarks on what the team has done, the results achieved and on what the team has yet to do.

Manuscript received May 14, 2007.

Paulo J. L. Adeodato and Germano C. Vasconcelos are Associate Professors at the Center for Informatics, Federal University of Pernambuco, Recife-PE, Brazil, (phone: +55 81 2126 8430; +55 81 2126 8438; e-mails: {pjla,gcv}@cin.ufpe.br).

Adrian L. Arnaud, Rodrigo C. L. V. Cunha, Domingos S. M. P. Monteiro are PhD Students at the Center for Informatics, Federal University of Pernambuco, Recife-PE, Brazil, (phone: +55 81 2126 8430; +55 81 2126 8438; e-mails: {ala2,rclvc,dsmpm}@cin.ufpe.br).

II. DATA SELECTION

As stated above, the approach presented here is robust particularly because it relies on sound statistical procedures.

Hence, a sample of the last 18 observations of each time series was separated only for performance assessment (the test set) aiming at preserving statistical independence from the parameter estimation process. The remainder of the data (the modeling set) was used for parameter setting and predictive modeling.

Trend analysis and elimination were carried out only considering the parameters extracted from the modeling set. The same was done for all the analyses conducted for lag definition and for establishing the parameters for data transformation, such as normalization and outlier filtering. Accordingly, the predictive modeling (learning) was also estimated only with the modeling data set.

III. DATA PREPARATION AND ANALYSIS

The modeling data was plotted for the 11 series problem for the modelers to get an intuitive feeling of the types of series present in the reduced problem. Several series had a repetitive yearly behavior, probably related to demand, few others had a very noisy trend while yet others had several noisy trends, probably related to financial variables.

Those with repetitive behavior had, in fact, their correlograms with a high spike on roughly the 12^{th} month lag. The latter had decreasing contributions from practically all lags.

The trends of the series were approximated by the best fitting linear functions which were subtracted from the series' data for later re-insertion, after the system prediction.

After trend removal, the series were normalized in such a way that all data points stayed in the range from 0.1 to 0.9 allowing for over 10% increase beyond these boundaries; 0-1 range.

These transformations were then applied to the remaining data set (the test set).

After these basic transformations, the temporal properties of the data series were checked through correlograms and Fourier analysis focusing on the time window needed for the predictive modeling. Since they (the correlograms and Fourier) produced different window sizes, both were considered on the optimization process by the modeling technique.

IV. PREDICTIVE MODELING

Considering the complexity of systematically modeling several time series with small amount of observations each, the best approach would be to develop a solution based on data mining with exogenous series for robustness and performance. However, this would be an expensive approach for the time available for the solution development.

The effective alternative was to develop robust solutions based only on each series data alone for the prediction of the future values. For this reason the solution relied on basic statistics over a number of independent trials: the median value predicted by 15 forecasting systems.

The well-known multilayer perceptron (MLP) was the modeling technique chosen. The MLP has been one of the neural network models most frequently used in pattern classification problems for its excellent generalization capacity, simplicity of operation and ability to perform universal function approximation [8]. It also presents robustness when compared to other techniques [9]. However, one drawback of this technique is the need of a validation data set for preventing over-fitting, which is critical in situations where there are few data observations available, such as in the case presented here.

The MLP chosen had a single hidden layer with processing units varying from 1 to 30 and was trained either with the standard error back propagation algorithm or with the Levenberg-Marquadt algorithm [10], having the minimum squared error on the validation set as the training stopping criterion.

If, in one side, the small amount of data was a drawback in terms of noisy solutions, on the other side, it allowed for a large amount of simulations for noise filtering through median forecasting.

Thus, each time series was exhaustively tested for both algorithms, all the possible number of processing units within the pre-defined range (1 to 30) and input window size (lags) based on either correlograms or Fourier analyses.

The architecture, training algorithm and input window selected were then replicated to produce 15 systems trained from a different initial state (weight initialization for symmetry breaking). The validation set was used as the training stopping criterion again and also for defining the temporal phase shift [11] in case it was identified to improve performance.

The test data set was then used for producing the forecasts for each one of the series by all the 15 systems.

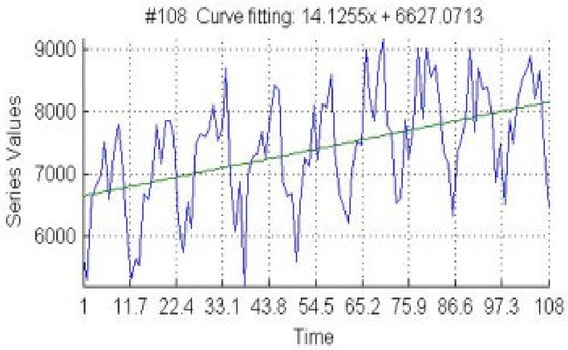
The 15 systems performance on both the validation and the test set were then compared for assessing two main aspects: quality degradation from the dependent to the independent data set and the quality itself, both measured in terms of the competition criterion: the SMAPE metrics. The choice was for the solution simultaneously closest to the median in terms of quality and to the median in terms of degradation.

V. RESULTS AND INTERPRETATION

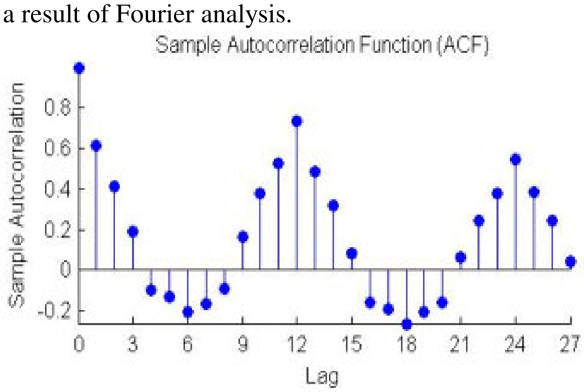
The results presented below show the graphs of some analyses carried out as well as the median series selected from the known data separated into validation and test sets, as described in the previous section.

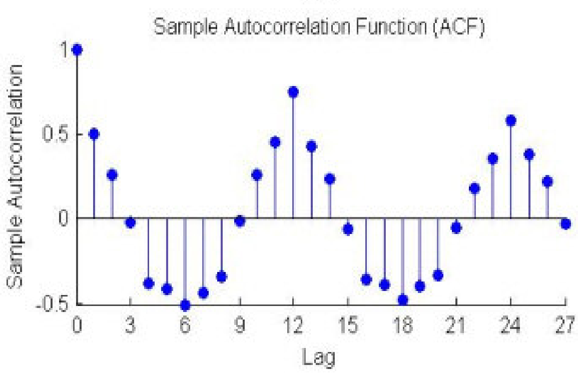
Two out of the 111 series were selected for illustrating the general ideas of this work because they represent the two most common types of behavior found on the problem: annual repetition and financial behavior. These series are the

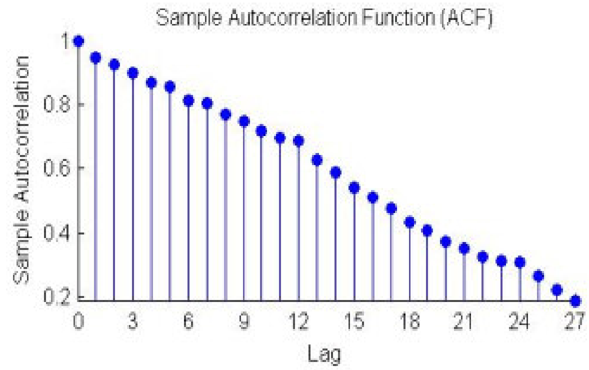
55th and the 74th from the 111 series set.

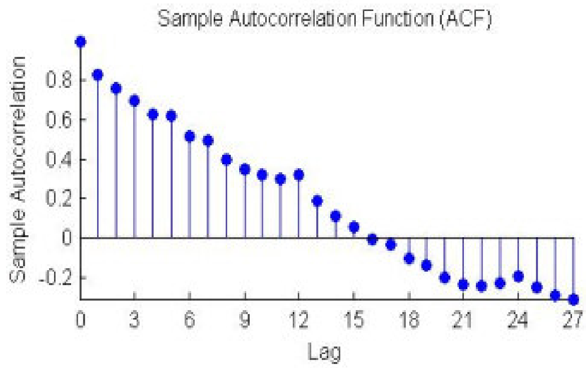


The correlograms of both series are show below before and after trend removal. They show clearly the behavior described above. A similar behavior has also been observed as a result of Fourier analysis.

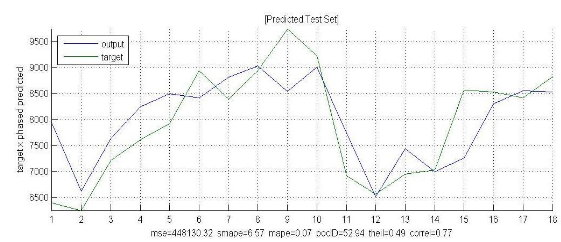


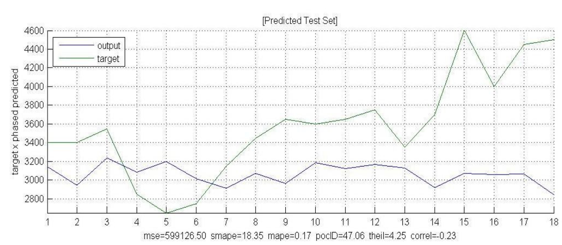






The following figures show for each time series the forecasts given by the elected solution, based on the median performance quality and degradation criteria described in the previous section. The comparison of the forecasts against the original data exhibits the difference in performance.





There is degradation between the forecasts provided for the test set when compared to the test set in both series. Degradation was worse for the financial series as one would expect from its correlogram; larger input window sizes would reduce the number of examples available for modeling.

The Tables below show the performance of the system comparing several metrics on the validation and test sets on both series. Some types of time series are affected by a phase shift when forecast by neural networks [11]. For this problem, phase correction was applied based only on the

SMAPE metrics measured on the validation data set. It was observed that for only few series (neither for #55 nor #74), the phase correction scheme generated performance improvement and had similar effects on the test set.

TABLE I
METRICS VALUES ON THE VALIDATION AND TEST SETS – SERIES 55

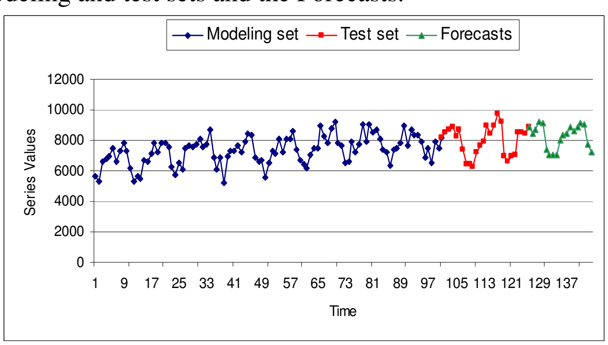
Metric	Validation set	Test set
MSE	120952.73	448130.32
SMAPE	3.45	6.57
MAPE	0.03	0.07
POCID	88.89	52.94
NMSE	0.17	0.49
CORREL	0.89	0.77

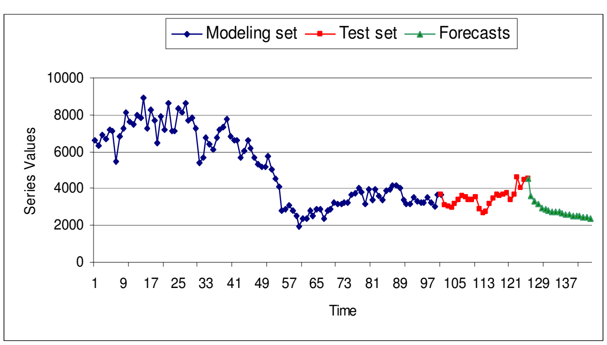
TABLE II
METRICS VALUES ON THE VALIDATION AND TEST SETS – SERIES 74

Metric	Validation set	Test set
MSE	64988.95	599126.50
SMAPE	6.31	18.35
MAPE	0.06	0.17
POCID	70.59	47.06
NMSE	0.78	4.25
CORREL	0.07	-0.23

For producing the forecasts the same methodology was applied to all series, this time taking the 18 last data observations as validation set for stopping training and for measuring all the correction factors on it. The 18 forecasts ahead produced by the 15 systems for the competition were then post processed and the median solution was selected to complete the submission file.

The following figures show the time series with the original data and the forecasts given by the elected solution. It shows the data separation in time intervals used as modeling and test sets and the Forecasts.





VI. CONCLUDING REMARKS

This paper has presented a principled methodology based on multilayer perceptron neural networks and basic statistics to produce robust time series forecasting.

A thorough procedure has been carried out for defining the most adequate combination of MLP architecture, training algorithm and time window for each series aiming at optimizing the SMAPE metrics. The robustness of the solution has been assured through the use of median forecasts of several systems.

Therefore, it is expected that the forecasts do not deviate from the actual values more than what had been measured during the modeling stage.

It is expected that the solution could be still further improved by the application of data mining for capturing relevant information from the exogenous series. Since this is an expensive approach it will be considered in a future work.

REFERENCES

[1] S. F. Crone, and K. Nikolopoulos, *NN3 Artificial Neural Networks & Computational Intelligence Forecasting Competition – Motivation*, (http://www.neural-forecasting-competition.com/motivation.htm), 2007.

[2] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. New Jersey: Prentice Hall, 1994.

[3] T. S. Rao and M. M. Gabr, *Introduction to Bispectral Analysis and Bilinear Time Series Models*, ser. Lecture Notes in Statistics. Berlin: Springer, 1984, vol. 24.

[4] T. Ozaki, *Nonlinear Time Series Models and Dynamical Systems, ser. Hand Books of Statistics*, E. J. Hannan and P. R. K. eds, Eds. Amsterdam: Noth-Holland, 1985, vol. 5.

[5] M. B. Priestley, *Non-Linear and Non-Stationary time Series Analysis*. London: Academic Press, 1988.

[6] 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.

[7] X. Yao, "Evolving artificial neural networks," in *Proceedings of IEEE*, vol. 87, no. 9, pp. 1423–1447, September 1999.

[8] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks*, vol. 2, pp. 359–366, 1989.

[9] M. Y. Kiang, "A comparative assessment of classification methods", *Decision Support Systems*, 35, pp 441-454, 2003.

[10] N. Ampazis, and S. J. Perantonis, "Two highly efficient second-order algorithms for training feedforward networks," *IEEE Transactions on Neural Networks*, Vol. 13, pp. 1064-1073, September, 2002.

[11] T. A. E. Ferreira, G. C. Vasconcelos, and Paulo J. L. Adeodato, "A New Evolutionary Approach for Time Series Forecasting", in *Proc. IEEE Symposium on Computational Intelligence and Data Mining - CIDM2007*, Island of Oahu – Hawaii, 2007.