16th Triennial World Congress, Prague, Czech Republic

SEASONAL TIME SERIES PREDICTION WITH ARTIFICIAL NEURAL NETWORKS AND LOCAL MEASURES

R. Pinto, S. Cavalieri

University of Bergamo, Department of Industrial Engineering, Italy
viale Marconi, 5. I-24044 Dalmine BG – Ph.: +39 035 2052385
E-mail: roberto.pinto@unibg.it

Abstract: Forecasting is one of the most challenging fields in the industrial research, due to its importance in practice and to the variability and number of elements that should be considered. In this context, the usage of Artificial Neural Networks has proved to be particularly advisable, thanks to their ability to approximate any kind of function within a desirable range. While most of the literature concerns with the definition of the characteristics of an ANN, there is only a few number of contributions that address the pre-analysis of the data, and seems there is no recent work about the pre-processing of the patterns to submit as input to the ANN. The aim of this paper is to propose a novel approach to the time series forecasting activities through the identification and exploitation of information hidden – or latent – into the values and the structure of a seasonal time series. In order to do this, the time series is decomposed into parts, for each of which some local measures are evaluated: such measures are intended to improve the forecasting ability of the ANN. Moreover, to exploit the "regularity" of a seasonal time series, the concept of *Seasonal Periodic Index (SPI)* has been introduced. Results obtained from the tests confirm the effectiveness of the usage of the local measures and of the *SPI*. *Copyright © 2005 IFAC*

Keywords: Artificial Neural Networks, Forecasting, Local measures, Seasonal time series, Levenberg-Marquardt learning algorithm,

1. INTRODUCTION

Forecasting is one of the most challenging fields in the industrial research, mainly due to its importance in any demand planning process and to the variability and number of elements that should be considered in the implementation of an effective prediction process, supported by suitable methods and techniques.
Forecasting techniques range from quantitative to subjective evaluation, depending on the objective of the forecast activity and the type of prediction being obtained.
Focusing on quantitative methods, two main classes can be singled out:

- *Extrapolative methods*: these methods aim to find some characteristics of a set of sequential observations (referred to as time series) that show a repetitive behaviour. Exponential smoothing techniques are an example of these methods.
- *Explicative methods*: these methods correlate one or more independent variables to one observed variable (i.e. dependent from the other variables), which has to be predicted. Examples are simple and multiple regression models.

In this context, Artificial Neural Networks (ANN), whose application in forecasting is discussed in this paper, are quantitative techniques which can be used either as universal regressors – thanks to their ability to approximate almost any kind of function (mainly, feedforward (Thiesing, and Vornberger, 1997; Crone, 2002) and Radial Basis Function networks (Ciocoiu, 1998; Zemouri, *et al.*, 2003) – or extrapolative methods, due to their feature to recognize pattern's characteristics through time (especially with time delay (Conway, *et al.*, 1998) and recurrent neural networks.

While most of the literature concern with the characteristics of an ANN (Hegazy and Salama, 1995) – i.e. topology, number of layers, number of neurons per layer, activation functions an so on – there are only a few contributions that address the pre-analysis of the data. For example, the analysis of the input data could be performed through the *Principal Component Analysis*, in order to reduce the size of the input space (Cloarec and Ringwood, 1998) and avoid the curse of dimensionality (the exponential growing calculation effort needed as the input space grows).

On the other hand, it seems that there is no recent work about the formation process of the patterns to submit as input to the ANN.

The idea at the basis of this work is that input patterns should be carefully organized in order to contain not only the values of the time series – as done generally in most applications – but also other information gathered from the time series itself.

This information could be obtained through an opportune pre-processing activity performed on the data.

In the following of this paper, an input pattern definition process is proposed, in order to enrich the information that the ANN could exploit for the forecasting.

More precisely, the paper is organized as follows: section 2 describes the differences between the traditional approach to forecasting with ANN and the proposed new approach; section 3 presents the assumption at the basis of the proposed forecasting model and the structure of the input pattern, characterized by the concept of local measures. In section 4 the definition process of the input pattern is presented: this process is the basis of the forecasting model, introducing the concept of Seasonal Period Index. Section 5 briefly exposes the structure of the ANN, while section 6 reports about the results obtained during the application of the model. Finally, in section 7 some conclusion remarks and further extensions are proposed.

2. FORECASTING TIME SERIES USING NEURAL NETWORK

The forecasting of a time series through an ANN could be performed in several ways, as stated in the previous section.

The general approach (with the exception of the usage of time delay ANNs) consists in the partitioning of the time series in a set of equally sized patterns – i.e. the same number of observations per pattern – that are presented to the ANN during the learning phase, together with the expected response.

This approach is based on the assumption that each observation A_i depends on the previous t values of the time series $A_{i-1}, A_{i-2}, \ldots, A_{i-t}$.

Starting from this approach, the idea at the basis of the model proposed in this work is to improve the performance of the ANN using information that is not explicitly represented by the time series values, but is hidden both in the values and in the "structure" of the time series.

In order to accomplish this, the assumption becomes more extended: each observation A_i of the time series could be seen as a combination of the *information* somewhat connected or represented by previous t observations.

It is important to note that we used the term *information*, which implies only an indirect reference to the numerical values of the observation, while referring strongly to other type of knowledge that is considered embedded but hidden in the time series.

Stated this assumption, the key of the forecast task becomes the identification and the exploitation of such information.

For example, the value of the next observation could be related to the mean of previous t values (as in the moving average methods): the mean of the t values is a piece of information which is different from the numerical value of the observations.

The reference to the moving average methods is significant also to explain another extension to the starting assumption, that seems to be reasonable: since the moving average considers only the last t values of the time series, we could assume that the information is *local*, i.e. is related only to a piece of the time series, and could be different depending on the part considered (see for example Nottingham and Cook, 2001)

For all these reasons, the proposed model, instead of focusing the attention on the structure and the characteristics that the ANN should have in order to perform adequately, focuses its attention on the data (i.e. the time series in input) and on the information that could be extrapolated from them, in order to support the forecasting activity.

In particular, the attention is concentrated on seasonal time series (with and without trend), because of the intrinsic difficulty involved in the prediction of such series.

3. THE ASSUMPTION OF THE FORECASTING MODEL

Each forecasting technique is based on some assumptions and hypothesis: our assumptions (or hypothesis) could be stated as follows:

a) Any behaviour of the time series observed in the past will repeat itself in the future.
b) Any observation of the time series could be seen as the result of the combination of the information embedded – implicitly or explicitly – in previous observations.
c) Such information is local to a particular subset of the available time series, i.e. such information changes depending on the time series under analysis.

d) There is a totally random fluctuation that could not be reasonably eliminated, but the entire series is not completely random (*Not Random Walk Hypothesis*)

The first assumption is the well-known *Hypothesis of Continuity*, which is the base for almost all the forecasting methods based on time series analysis: we could only predict the future studying the past and assuming that the past will repeat itself in the future, hopefully almost unvaried.
The second assumption is derived from the so called *Weak Form Efficient Market Theory* (Fama, 1970; Sitte and Sitte, 2002) which the Random Walk model (mentioned in the fourth assumption) is an extension of. As explained in section 2, each observation is seen not only as a combination of previous values (such as in the moving average methods or in the exponential smoothing technique), but it is also influenced by the information (i.e. not explicit, but in some case easily obtainable) embedded into previous values. Such information is quantitative, due to the mathematical nature of the ANN.
The third assumption states that, if we consider only a subset of the time series, we could find in such a set enough information to be used to forecast the next value of the series. In other words, each observation has in its neighborhood a hidden, latent information that could help in the prediction.
Finally, the fourth assumption says that it is almost impossible to predict exactly the values in a real time series, because of the presence of random and unknown hidden phenomena. It is difficult to model or quantify such elements, so the prediction should not be represented by a single value, but by a range of possible values.
Given these assumptions, in our work we attempted to improve the forecasting results provided by a feedforward ANN by exploiting information embedded into the time series. More precisely, we would say local information, meaning a set of measures (as said, the information considered is quantitative) evaluated considering only a few observations at a time.
In practice, time series are partitioned into subsets, each characterised by the following measures: (i) mean value; (ii) variance; (iii) slope of the regression line that fits the subset (referred to as regression slope in the following).
Each subset, with its related local measures, is the basis for the construction of the input patterns for the network learning process.

4. DEFINITION OF THE INPUT PATTERN

The main task of the proposed model is the definition of the input patterns of the ANN, starting from the sequence of values of the time series.
Before describing the structure of the input patterns, it is useful to introduce some definitions:

- *LS*: the length of the time series (i.e. the number of the observations in the series);
- *W*: the "window extension", meaning the number of time series observations that constitute an input pattern;
- *L*: the number of the observations within a single seasonal cycle; for example, if the season is over one year and the observations are monthly registered, L will be equal to 12; if, instead, the observations are gathered every 3 months, L will be equal to 4.

Starting from a seasonal time series of length *LS*, each input pattern will be formed by *W* sequential observations, while the target value is represented by the *W+1* value.
As defined previously, we make use of local information for each pattern, represented by the mean and the variance of the *W* observations of the patterns, plus the slope of the regression line that fits such *W* observations.
Moreover, in order to exploit the "regularity" of seasonal time series, we introduce another element to the discussion: if the time series has a seasonality period of *L* observations (i.e. the series repeats itself every *L* observations) we could provide the ANN with such information, thus reducing the training period and improving the quality of results.
For this reason, beyond the above illustrated measures, we have added a further set of inputs, namely the *Seasonal Period Index (SPI)*. This is an array of *L* binary values, which is subjected to the following constraint:

$$\sum_{i=1}^{L} SPI[i] = 1$$

where *SPI[i]* represents the seasonal array.
Thus, for each pattern only one element of the SPI could be set to 1. The position *p* of the array which stores the non-zero value in a pattern is defined according to the following expression:

$$p = h \bmod(L) + 1$$

where *h* is the position of the target value in the time series for the forming pattern (so $h \in [1 ; LS]$). Obviously, $p \in [1 ; L]$.
As a result, we have a set of patterns composed by *W+3+L* elements.
Figure 1 may help in understanding the process illustrated above.
As shown, the first pattern is composed by three sequential values of the time series (*W* = 3). The mean, the variance and the regression slope of these three values are computed and compose the other three elements of the pattern.
Most interesting is the last part of the pattern: supposing that the season is over four periods, we add four binary values as defined before.

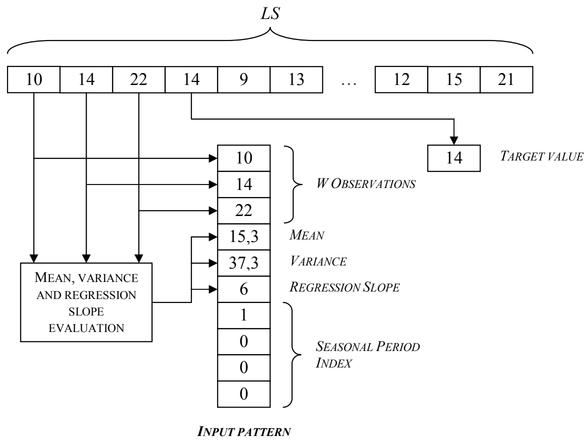

Fig. 1. The input pattern definition process

In order to define which is the only non-zero value, we have to look at the target value: since the target value is in the fourth position ($h = 4$) and $L = 4$, the non-zero position in the *SPI* is:

$$P = 4 \bmod(4) + 1 = 1$$

So, the 1 should be positioned in the first cell of the *SPI* array. The second pattern will have the 1 in the second position of the *SPI* array and so on.

5. THE ANN AND THE LEARNING ALGORITHM

The ANN used in this work is a feedforward backpropagation network with one hidden layer (with logistics activation function) trained with the Levenberg-Marquardt (LM) supervised learning algorithm.
This algorithm, as the quasi-Newton methods, has been purposely designed for providing an approximation of the Hessian matrix of the performance function, improving the speed of learning, especially for moderate-sized networks.
This algorithm revealed to be the best one in the test we conducted, both in the epochs needed for reaching the target error and in the quality of the result, measured in terms of *Mean Average Percentage Errore (MAPE)*:

$$MAPE = \sum_i \frac{|o_i - t_i|}{t_i} \cdot 100$$

where o_i is the output of the network and t_i is the target value, that is the desired value.
Moreover, the LM algorithm is particularly suitable whenever a relatively small number of observations is available.
Two different objective functions have been tested: the simple *Mean Squared Error* (*MSE*) and the *MSE* with regularization (*MSEREG*), as defined in the Matlab™ package:

$$MSEREG = \lambda \cdot MSE + (1 - \lambda) \cdot MSW$$

$$MSW = \frac{1}{n} \cdot \sum_{i=1}^{n} w_i^2$$

where λ is called performance ratio ($\lambda \in [0\,;\,1]$; if $\lambda = 1$, then $MSEREG \equiv MSE$) and w_i is the i-th weight of the ANN.
This function, used as the objective function of the learning phase, aims to reduce the weight of the network, forcing the network response to be smoother and avoiding the risk of data overfitting, that occurs if the ANN has memorized well the patterns used for the training, but shows really poor performances on new data sets.

6. RESULTS

The proposed approach has been tested over a number of seasonal sales time series, obtained from various sources.
For the sake of clarity, only the test conducted over a seasonal time series with trend (Figure 2) is reported.
Tests have been conducted over an ANN with a hidden sigmoidal layer of a variable number of neurons (ranging from 5 to 10), trained with the same learning algorithm (the LM) but with different inputs:

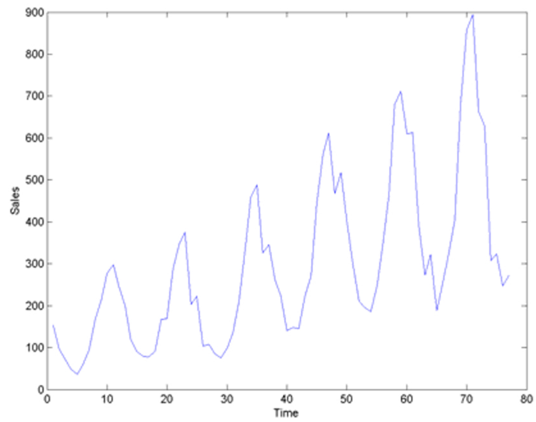

Fig. 2. Sample of the data used for the test

A) with the set of pattern constituted from the rolling *W* observation of the time series;

B) with the set of pattern including the *W* rolling observation, the local measures and the *SPI*.

The results of a series of test are reported in Table 1, where NN_LM indicates an ANN trained with the patterns using the local measures and *SPI*, constructed as illustrated in section 4.
Each network and pattern configuration has been tested 7 times, and Table 1 reports the mean, the minimum and the maximum MAPE reached both on training and test set.
As could be seen, local regressions measures improve the response of ANNs in almost all the tests conducted.
Moreover, the usage of the *MSEREG* function results in an improvement of the generalization ability of the network.

Table 1 Results of the test

	Training Set		Test Set			
	NN_LM	NN	NN_LM	NN	Window Size	Objective
Mean Error (%)	9,42	17,63	13,91	25,65	5	MSEREG
Min Error (%)	6,79	8,90	10,51	16,97		
Max Error (%)	11,00	30,68	16,44	49,15		
Mean Error(%)	10,83	17,46	16,69	24,34	6	
Min Error (%)	9,37	13,65	14,42	18,62		
Max Error (%)	15,34	21,51	24,28	31,42		
Mean Error (%)	9,77	16,93	15,39	24,34	7	
Min Error (%)	6,25	11,95	9,79	15,14		
Max Error (%)	14,18	23,80	22,89	36,56		
Mean Error (%)	10,06	32,02	15,90	54,35	5	MSE
Min Error (%)	7,88	20,87	11,76	34,73		
Max Error (%)	11,89	38,98	19,51	66,73		
Mean Error (%)	11,51	27,76	18,34	46,82	6	
Min Error (%)	7,80	19,26	11,55	32,15		
Max Error (%)	20,26	42,46	33,74	72,74		
Mean Error (%)	13,48	21,71	22,13	36,15	7	
Min Error (%)	8,35	11,11	13,13	17,49		
Max Error (%)	21,38	34,27	36,27	58,31		

Figure 3 reports more clearly the differences between the two approaches: with the usage of local measures and *SPI* the forecast on the training set is more accurate.

As stated before, ANN could approximate arbitrarily well almost any kind of function, given a suitable set of data.

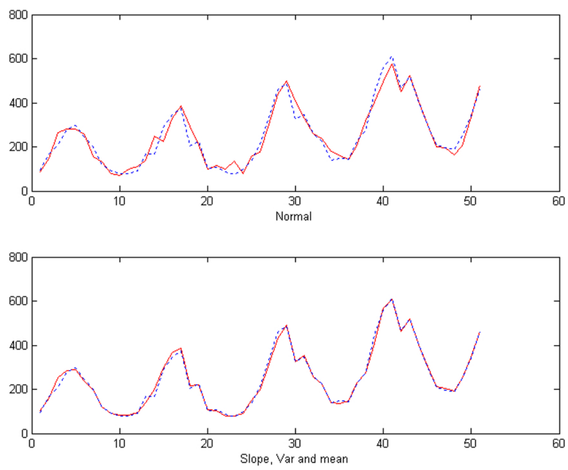

Fig. 3. Time series (dotted line) versus forecast on the training set

Most of the problems arise when the ANN has to *generalize* its knowledge, i.e. when the ANN has to predict the next value of a time series on the basis of inputs never seen before. The usage of the *MSEREG* avoids the overfitting problem, but nothing could be said precisely about the performance of the network when tested on a set of inputs that have not been presented to the ANN during the training phase.

This is the reason why a *test set* has also been used: the test set is made up by observations that have not been presented to the ANN during the training process.

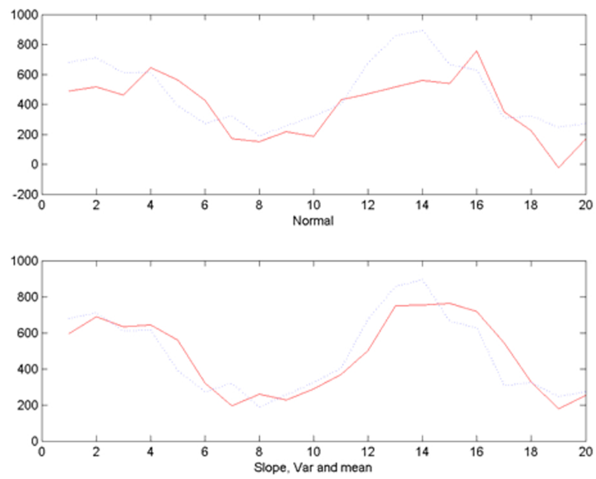

Fig. 4. Time series (dotted line) versus forecast on test set

As could be seen in Figure 4, that depicts the forecast on the test set, the proposed model outperforms the normal pattern network also on the test set.

Analysing the results of the test, the following evidences emerge:

- *The assumptions of the model are validated:* the assumptions stated in section 3 has revealed to be reasonable; that is, in seasonal time series could be found a latent information that could improve the forecasting process.
- *The local measures are appropriate:* in the proposed model, the mean, variance and regression slope are adopted. Although these measures seem suitable to improve the forecast, there could be other local measures that could be exploited.
- *The usage of Seasonal Period Indicator improves the performances:* providing the information about the period to which the forecast will belong contributes to improve the performances of the ANN.
- *A simple network architecture works fine:* although it is possible to use different typologies of ANN, the tests confirm that also a relatively simple feedforward neural network can show good performances.
- *The generalization ability of the network is improved:* perhaps the most important result, the added information provided by local measures and SPI improves the forecasting ability on new data.
- *The model is robust against trend:* this result emerges from the comparison of the result swith the application of de-seasonalizing methods: such methods, in fact, could perform well only if the time series is stable, while the proposed method works fine also in presence of trend (as in the example of the time series depicted in Figure 2)

7. CONCLUSION: REMARKS AND FURTHER EXTENSIONS

In this work, a novel approach to the usage of ANN for forecasting has been proposed. Instead of focusing the attention on the characteristics of the ANN, we have proposed a model which aims to better exploit the information embedded into the historical data.

The results prove that the usage of so called local measures (like mean, variance and slope regression) could improve significantly the performance of an ANN. Moreover, the introduction of the *Seasonal Period Index* (*SPI*) contributes to the effectiveness of the model.

In this work, only a part of the possible measures and pattern configuration has been examined.

In order to generalize the results, some of the most challenging extensions of the present work could be the following:

- the identification of the suitable structure of the ANN (in particular, the number of the neurons of the hidden layer) by using some correlation measures of the characteristics of the time series itself, in order to reduce the size of the network and the training epochs;
- although the feedforward network with LM algorithm performs well, could be interesting to extend the analysis to other network architectures, such as the *Generalized Regression Neural Networks (GRNN)* often used for function approximation;
- another interesting topic is the definition of the extension of the window *W* to consider in the transformation of the time series into input pattern. This parameter could be linked, for example, to the seasonality of the time series or to other characteristics;
- finally, the proposed method could be extended to other non-seasonal time series, that could represent one of the most attractive fields of future research.

REFERENCES

Ciocoiu, J.B., (1998). Time series analysis using RBF networks with FIR/IIR synapses. *Neurocomputing*, **20**, 57 – 66

Cloarec, G.M., Ringwood, J., (1998). Incorporation of Statistical Methods in Multi-step Neural Network Prediction Models. 0-7803-4859-1/9, IEEE – 2513-2518

Conway, A.J., Macpherson, K.P., Brown, J.C., (1998). Delayed time series predictions with neural networks. *Neurocomputing*, **18**, 81 – 89

Crone, S.F., (2002). Training artificial neural networks for time series prediction using asymmetric cost functions. Technical Working Paper IWI-0201, accepted at: ICONIP 2002 Proceedings

Fama, E.F., (1970). Efficient capital market: A review of theory and empirical work. *Journal of Finance*, **25**, 383 – 417

Hegazy, Y.G., Salama, M.M.A., (1995). An Efficient Algorithm for Identifying the Structure of Artificial Neural Networks for Forecasting Problems. 0-7803-2428-5/95, IEEE – 618-621

Nottingham, Q.J., Cook, D.F., (2001). Local linear regression for estimating time series data. *Computational Statistic & Data Analysis*, **37**, 209-217

Sitte, J., Sitte, R., (2002). Neural Networks Approach to the Random Walk Dilemma of Financial Time Series. *Applied Intelligence*, **16**, 163 – 171

Thiesing, F.M., Vornberger, O., (1997). Sales forecasting Using Neural Networks. Proceedings ICNN '97, Huston, Texas, 9-12 June 1997, vol. 4, pp. 2125 – 2128, IEEE

Zemouri, R., Racoceanu, D., Zerhouni, N., (2003). Recurrent radial basis function network for time-series prediction. *Engineering Application of Artificial Intelligence*, **16**, 453 – 463

http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/