ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 93-102

Soft-computing techniques for time series forecasting

I.Rojas, H.Pomares

Department of Computer Architecture and Computer Technology
E.T.S.I.I. University of Granada (Spain)

Abstract

One way to contrast the behaviour of different algorithms in the field of time-series forecasting is to compare the prediction error using a benchmark problem. Another interesting way is to perform a competition. In this paper we shortly discuss the competition organized by EUNITE for an electricity load forecasting. Given the temperature and the electricity load from 1997 to 1998, the competitors are asked to supply the prediction of maximum daily values of electrical loads for January 1999 (31 data values altogether, including some holidays). In total, 56 registered competitors from 21 countries were submitted. A summary of the contribution of the best papers along with some remarks about the use of soft-computing in time-series forecasting and future trends are presented in this paper.

1. Introduction

The challenge of predicting future values of a time series spans a variety of disciplines. The multiplicity of techniques developed to make predictions manifests a heritage from biology, computer science, economics, engineering, mathematics, physics, statistics, and other areas. These methods find application in such diverse data sets as animal populations, equity market prices, disease control, meteorological measurements, astronomic observations, and others.

Time series analysis refers to the branch of statistics where observations are collected sequentially in time, usually but not necessarily at equally-spaced time points, and the analysis relies, at least in part, on understanding or exploiting the dependence among the observations. The goal of time series prediction can be stated succinctly as follows: given a sequence up to time t x(1), x(2),..., x(t), find the continuation x(t+1), x(t+2), ...

Because of the importance of time series analysis, many works in the literature can be found about this subject, especially those based on statistical models. There are many possible techniques for capturing temporal behaviour. Box-Jenkins's ARIMA (Autoregressive Integrated Moving Average) time series analysis is attractive [3] because it provides a comprehensive statistical modelling methodology for I/O processes. It covers a wide variety of patterns, ranging from stationary to non-stationary and seasonal (periodic) time-series.

Briefly, Box and Jenkins [3] described a general linear stochastic model as one that produces output whose input is white noise ε_t, or a weighted sum of historical ε_t. Mathematically, it can be expressed as follows:

$$\widetilde{Y}_t = \mu + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \ldots - \varphi_q \varepsilon_{t-q_t} - \ldots \quad (1)$$

where, μ is the mean of a stationary process, and φ_t, t = 1,2,..., are coefficients which satisfy

$$\sum_{i=0}^{\infty} \varphi_i^2 < \infty \quad (2)$$

ε_t is an uncorrelated random variable with mean zero and constant variance σ_ε^2. However, equation (1) is not very practical because it involves an infinite number of terms. It is thus convenient to express equation (1) in terms of a finite number of autoregressive (AR) and/or moving average (MA) components. If Y_t is defined as $\widetilde{Y}_t - \mu$, the deviation of the process from some origin, or from its mean, an AR(*p)* process can be generally expressed as follows:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \quad (3)$$

Likewise, an MA(*q)* process can be expressed as follows:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q_t} \quad (4)$$

Thus, a general express ion for a mixed ARMA(*p, q)* process can be defined as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q_t} \quad (5)$$

where ϕ_t and θ_t are coefficients that satisfy stationarity and invertibility conditions, respectively.

One problem that plagues the development and implementation of these time series models is the requirement of a formal model specification with an assumed probability distribution for the data [11]. In order to address this problem, McDonald and Xu (1994) [12] apply partially adaptive estimation techniques that use very flexible families of distributions. Their approach allows varying degrees of skewness and kurtosis in time series data. However, Box–Jenkins methodology is some-times inadequate for situations where the underlying failure behaviour varies dynamically with time. Hence, other approaches based on nonlinear techniques such neural network, fuzzy logic, evolutionary algorithm, SVM, etc, should be analyzed [4].

Artificial Neural networks (ANNs) offer some potential advantages over ARIMA time series models in dealing with non normal and nonlinear data problems. One of the advantages of ANNs is their versatility and that they do not require formal model specifications or assumed probability distributions for the data. Another important issue is that, as it is shown in [13], ANNs are capable of tolerating chaotic components having very thick tails better than most alternative methods. Because many important time series have significant chaotic components, this capability is important. Hill et al [11] used 104 time-series data sets from the same M-Competition and compared the performance of feedforward neural networks with that of six other traditional statistical models including Box –Jenkins ARMA model. They found that neural networks performed significantly better than traditional methods for monthly and quarterly time series. For annual time series, however, Box –Jenkins' model seemed to be comparable to neural networks.

Although most of the studies reviewed above indicate neural-network models are comparable or superior to linear models for particular data series, it is questionable whether or not neural networks can consistently outperform the Box–Jenkins models in all situations. Some authors have observed that it may be possible that the neural-network approach will outperform some standard forecasting procedures such as the Box –Jenkins model [15][20], but that it may only do so for certain situations. Some reviews even argue that there is little evidence as yet that ANNs might outperform standard forecasting methods [14].

2. Eunite competition: Electricity Load Forecast using Intelligent Adaptive Technology

The time series competition data was supplied by the Eastern Slovakian Electricity Corporation (http://www.vse.sk/) that claim an important interest of the application of Intelligent and adaptive technologies in electricity load forecast. This is very important issue, which can bring a very significant financial profit using more accurate prediction technology. In order to forecast the maximum daily values of electrical loads for January 1999, the following data was supplied (http://neuron.tuke.sk/competition/ index.php):

Load1997.xls	- the file consists of all half an hour electrical loads of the year 1997. Every line represents a corresponding day (from 1 to 365) and every column presents half an hour loads (from 1 to 48).
Load1998.xls	- the file consists of all half an hour electrical loads of the year 1998. Every line represents a corresponding day (from 1 to 365) and every column presents half an hour loads (from 1 to 48).
Temperature1997.xls	- the file consists of average daily temperatures of the year 1997.
Temperature1998.xls	- the file consists of average daily temperatures of the year 1998.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 93-102

Holidays.xls	- the file describes the occurrence of holidays in the period from 1997 – 1999.

The result of each solution was evaluated according to the following two indexes:

$$\mathrm{MAPE} = 100 \cdot \frac{\sum_{i=1}^{31} \left| \frac{Load_{\mathrm{Real}}(i) - Load_{\mathrm{Predicted}}(i)}{Load_{\mathrm{Real}}(i)} \right|}{31} \tag{6}$$

$$\mathrm{MAXIMAL\,error\,(M)} = \max\left(\left| Load_{\mathrm{Real}}(i) - Load_{\mathrm{Predicted}}(i) \right| \right), i = 1,\ldots 31$$

It is important to note, that in this competition a long-term prediction is required, which is much harder and difficult than a short-term prediction due to the error propagation. In fact, in order to forecast the whole month, also exogenous variables should be taken into account such as the temperature values and the information about the holidays throughout the target month.

In order to gain an insight on the specific problem we are dealing with, in figure 1a shows the electricity load for the two previous years arranged sequentially in time. Figure 1b zooms in the behaviour of the load for a specific day.

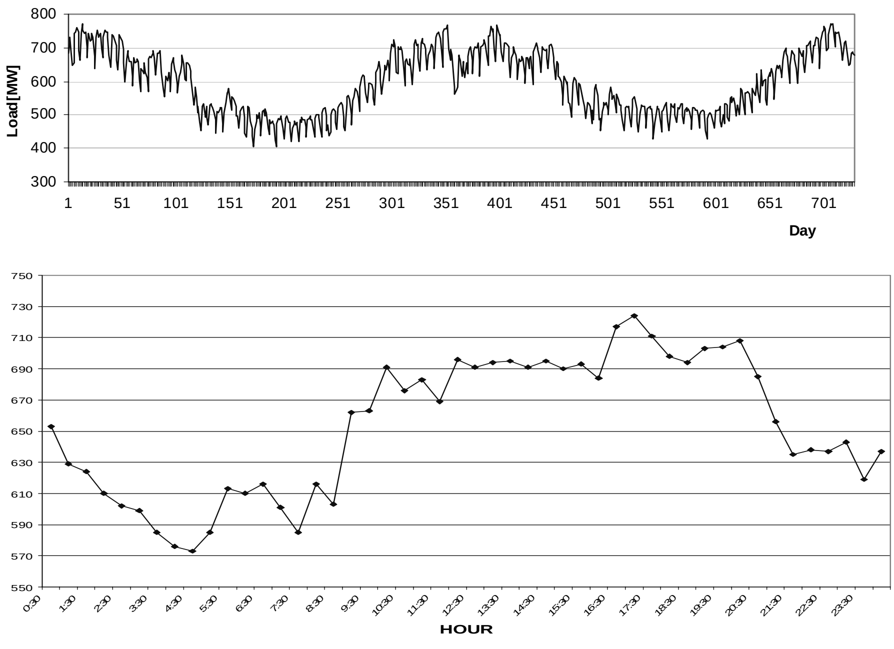

Figure 1: a) Maximum electricity load for every day in years 1997 and 1998. b) Electricity load pattern for a specific day.

In total, 56 registered competitors from 21 countries were submitted. For each of the competitors, the MAPE and the Maximal Error was computed using as the test set, the real data obtained for January 1999 (which are plotted in figure 2). Figure 3 shows the prediction made by the three papers that win the competition and which we shortly describe in the next section.

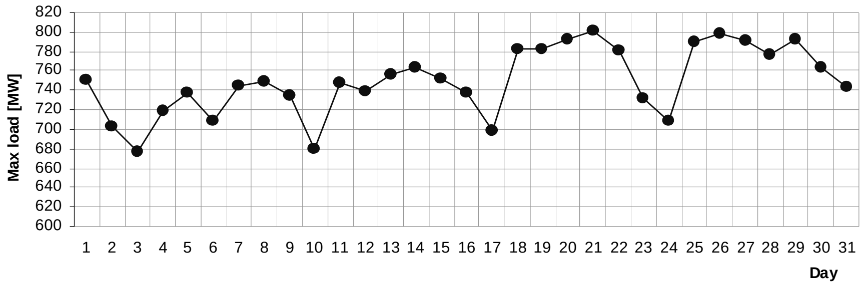

Figure 2: Real data of the maximum load for every day in January 1999

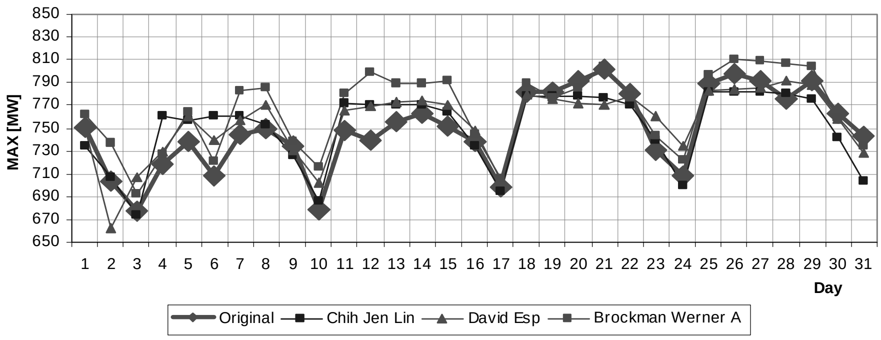

Figure 3: Original test set and predicted data obtained by the three winners of the competition

3. Best methodologies for electricity load forecast

The three best methodologies proposed for the EUNITE competition were those of:

- EUNITE Network Competition: Electricity Load Forecasting, by M-W. Chang, B-J Chen and Chih-Jen Lin [5]
- Adaptive Logic Networks for East Slovakian Electrical Load Forecasting, by David Esp [6]
- Different models to forecast electricity loads, by W.Brockmann and Steffen Kuthe [7]

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 93-102

3.1 Winning paper

The work proposed by Chang et al, was based on Support Vector Machines (SVM) [9], [10]. Given a set of data points (x_1,y_1), ..., (x_ℓ,y_ℓ), where x_i are the input and y_i the scalar output values, SVM approximates this data set by the following expression:

$$f(x) = w^T \cdot \phi(x) + b \qquad (7)$$

where $\phi(x)$ represents the high-dimensional feature space and w and b are estimated by minimizing the following regularized risk function:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} w^T w + C \sum_{i=1}^{\ell} \left(\xi_i + \xi_i^*\right) \qquad (8)$$

subject to

$$\begin{aligned} & y_i - \left(w^T \phi(x_i) + b\right) \le \varepsilon + \xi_i \\ & \left(w^T \phi(x_i) + b\right) - y_i \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0, \quad i = 1..\ell \end{aligned} \qquad (9)$$

After doing some analysis to the data provided by the organization they tried to estimate the electricity load using as inputs:

- Seven attributes for maximal loads of the past seven days
- Seven binary attributes for the day of the week
- One binary attribute indicating whether it is a holiday or not
- One attribute indicating the daily average temperature

As training data, they used data for 1997 and 1998 excluding data from abril to september and january 1998. As validation data: data for january 1998.

Nevertheless, they realized that forecasting the daily average temperature is a problem as complex as the one they were analysing (since this is a datum that was not given by the competition organizers). So, after some trial and error, the finally conclude their investigation discarding the information of the temperature (and also that provided by the holiday information).

Finally, for their SVM algorithm, they used RBF functions with ε, C and γ obtained by cross-validation. Their winning MAPE performance over the unseen test data was: 1.98% and the maximum prediction error was 51.4 units.

3.2 Second and Third position

Now we briefly describe the proposed methodologies of the second and third award papers.

In [6], David Esp used a modelling technique based on the Adaptive Logic Network (ALN) which is a continuous value generalization of a Boolean modelling technique originated by W.Armstrong in 1968 [8]. The main idea behind the ALN learning algorithm is least-squares fitting, starting off with plain linear regression, then progressively splitting into a multitude of separate partial linear fits.

For his experiments, he used the electricity load of January 1996, data which were not available at first for the participants but were given to the author afterwards. After some initial analysis the author opted to use the data from January and February only, discarding the data from any other months. The influence of the different parameters of the ALN was studied by trial and error. The final "best" method was to predict the electricity loads for each day (every half an hour) and take the maximum using as inputs:

- day of the year
- time of day
- temperature
- if the day is a holiday, Saturday or Sunday

The main problem the author encountered was to estimate the temperature. For that purpose, he averaged the temperature for a given day from that of the same day of the previous years. The obtained MAPE performance over the unseen test data was: 2.14% and the maximum prediction error was 40.0 units.

Finally, the third best result was that obtained by W.Brockmann and Steffen Kuthe [7]. They proposed two different methodologies. One based on simple statistical theory and the other based on a mix between parametric modelling and neuro-fuzzy systems. For the latter, they propose to model the series decomposing it in a base load plus a seasonal term plus a weekly term. The output of that model is afterwards corrected by a neuro-fuzzy system whose inputs are the day of the week and a holiday indicator. For the former approach, the simplest one and which gave them the third position in the competition, they calculate the average values between January 1997 and January 1998 to predict the values for January 1999. In order to correct the effect of different week-days, the shift the data sets of year 1997 for two days and of 1998 for one day, such that Mondays relate to Mondays and so forth. In this case, the MAPE obtained was 2.49% and the maximum prediction error 60.5 units.

4. Future trend and conclusion

The evolution of the number of papers, per year of publication, obtained by IEEE Xplore (http://ieeexplore.ieee.org/Xplore/) in IEEE Journal and Conference proceedings is shown in Figure 4. It is important to note, that, of course, this is just an example, there ara other discipline or set of journal as Kluwer Journal, ScienceDirect, Springer, etc. Nevertheless, we will try to explore the evolution of publication related with "time series" and soft computing techniques, and also with the wining paper of the EUNITE competition that was based on Support Vector Machines (SVM).

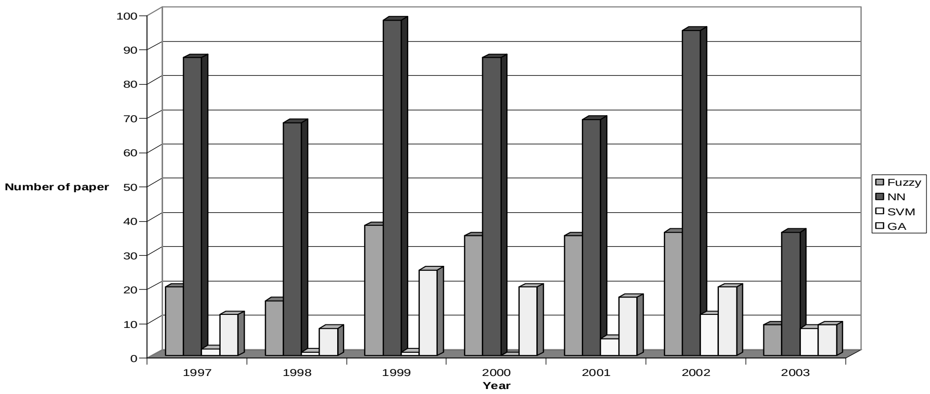

Figure 4: Evolution of the number of papers on different field of soft-computing in time series analysis published in IEEE Xplore

It is important to highlight the evolution of paper focussed in SVM during 2001-2003, taking into account that the EUNITE competition was in 2001.

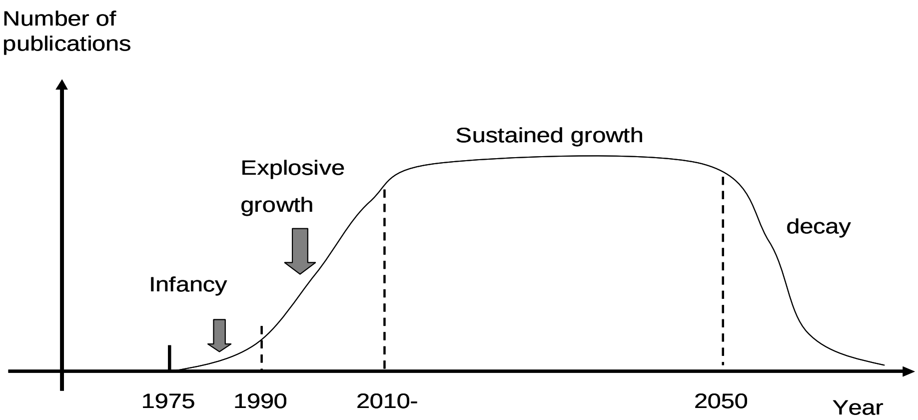

Figure 5: Predictable development of the applications of soft-computing techniques in time series analysis

It is difficult to predict the future, but considering the classic shape representing the life of the scientific disciplines (Figure 5), the field of time series analysis using soft-computing techniques has left behind the initial period of slow and moderate growth and is now in the phase of strong growth. One can expect that this phase will last until 2010, then will reach the plateu for about 10-15 years (depending of the new descoveries).

Finally, we want to mention the Time Series Prediction Competition that will be celebrated during the International Joint Conference on Neural Networks, July 2004. The proposed time series is the CATS benchmark (for Competition on Artificial Time

Series) has 5,000 data. Within those 100 values are missing. These missing values are divided in 5 blocks: elements [981 to 1,000], [1,981 to 2,000], [2,981 to 3,000], [3,981 to 4,000] and [4,981 to 5,000]. These 100 missing values have to be predicted.

Acknowledgement
This work has been partially supported by the CICYT Spanish Project DPI2001-3219.

References

[1] Hill T.,O 'Connor M.,Remus W. "Neural network models for time series forecasts". Management Science 1996;42:1082 –92.
[2] Kohzadi N,Boyd M.S, Kermanshahi B, Kaastra I., "A comparison of artificial neural network and time series models for forecasting commodity prices", Neurocomputing 1996;10:169 –81.
[3] Box G.E.P, Jenkins G.M, Reinsel G.C. "Time series analysis: forecasting and control", Englewood Clis,NJ:Prentice-Hall, 1994.
[4] K. Xu, M. Xie, L.C. Tang, S.L. Ho, "Application of neural networks in forecasting engine systems reliability", Applied Soft Computing 2 (2003) 255–268.
[5] M-W. Chang, B-J Chen, C-J Lin, "EUNITE Network Competition: Electricity Load Forecasting", November 2001. Winner of EUNITE world wide competition on electricity load prediction.
[6] D. Esp, "Adaptive Logic Networks for East Slovakian Electrical Load Forecasting", EUNITE World-wide competition, November 2001.
[7] W.Brockmann, Steffen Kuthe, "Different models to forecast electricity loads", EUNITE World-wide competition, November 2001.
[8] W.Armstrong, "On networks of Adaptive Boolean Logic Elements and their Application to Pattern Recognition", Proc. 4th Princeton Conference, 1968.
[9] V.N.Vapnik, "The Nature of Statistical Learning Theory", New York, Springer-Verlag, 1995.
[10] B. Scholkopf, A.J. Smola, "Learning with Kernels", MIT Press, Cambridge, MA, 2002.
[11] J.V. Hansen, J. B. McDonald, R. D. Nelson, Time Series Prediction With Genetic-Algorithm Designed Neural Networks: An Empirical Comparison With Modern Statistical Models, Computational Intelligence, v.15, no.3, 1999.
[12] J.McDonald, Y.Xu, "Some forecasting applications of partially adaptive estimators of ARIMA models", Economics Letters, vol.45, pp.155-160,1994.
[13] T.Masters, "Advanced algorithms for Neural Networks", John Wiley and Sons, New York, 1995.
[14] J. Nowicka-Zagrajeka, R. Weronb, "Modeling electricity loads in California: ARMA models with hyperbolic noise", Signal Processing 82, pp.1903–1915, 2002.
[15] Wei Yee Goh; Chee Peng Lim; Kok Khiang Peh, "Predicting drug dissolution profiles with an ensemble of boosted neural networks: a time series approach", IEEE Transactions on Neural Networks, vol.14 , n.2, pp.

459 – 463, March 2003.

[16] Povinelli, R.J.; Xin Feng, "A new temporal pattern identification method for characterization and prediction of complex time series events" IEEE Transactions on Knowledge and Data Engineering, vol.15,n.2, pp. 339 – 352, March-April 2003.

[17] Versaci, M.; Morabito, F.C., " Fuzzy time series approach for disruption prediction in Tokamak reactors" IEEE Transactions on Magnetics, vol.39, Isssue, 3,pp.1503 - 1506 May 2003

[18] Berardi, V.L.; Zhang, G.P.;" An empirical investigation of bias and variance in time series forecasting: modeling considerations and error evaluation" IEEE Transactions on Neural Networks, vol.14 , n.3, pp. 668 – 679, May 2003

[19] Castillo, O.; Melin, P.;" Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory" IEEE Transactions on Neural Networks, vol.13 , n.6, pp. 1395 – 1408, Nov. 2002

[20] Kasabov, N.K.; Qun Song; "DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction" IEEE Transactions on Fuzzy Systems, vol.10, n. 2, pp. 144 – 154, April 2002