Medium Term Electric Load Forecasting Using TLFN Neural Networks

Danilo Bassi, Oscar Olivares

Abstract: This paper develops medium term electric load forecasting using neural networks, based on historical series of electric load, economic and demographic variables. The neural network chosen for this work is the Time Lagged Feedforward Network (TLFN), which combines conventional network topology (multilayer perceptron) with good handling of time dependencies by means of gamma memory. This is a versatile mechanism that generalizes the short term structures of memory, based on delays and recurrences. This scheme allows smaller adjustments without requiring changes in the general network structure. The neural model gave satisfactory results exceeding those obtained by classical statistical models like multiple linear regression.

Keywords: Neural network model, forecasting, gamma memory, electric load

1 Introduction

Electric load is one of the key variables for electric power companies, since it determines its main source of income, particularly in the case of distributors. According to the foreseen load, the company makes investments and decisions on buying energy from the generating companies, and planning for maintenance and expansion. It is therefore absolutely necessary to have some knowledge of future power load: electric power distributors require a tool that allows them to predict the load in order to support its management and make more efficient its planning formulation. Accurate prediction of electric load is difficult, because it is determined largely by variables that involve "uncertainty" and whose relation with the final load is not deduced directly. The load is also characterized as a nonlinear and nonstationary process that can undergo rapid changes due to weather, seasonal and macroeconomic variations. A large number of the classical prediction models are inappropriate for this modelling because of their requirement of linearity and seasonality. On the other hand, an important contribution of Artificial Neuronal Networks (ANN) is their elegant capacity for approaching arbitrary nonlinear functions, which makes them adequate for solving this kind of problem. This paper proposes a solution to the problem of predicting monthly electric loads (medium term) of low voltage and low installed power (BT1) customers of distributing companies in Chile, based on both historical load values and on the country's economic and demographic growth data. To solve this forecasting problem various neuronal network models were analyzed and applied to a real case with information from four electric distribution companies in the north of Chile. Different methodologies for forecasting time series in various areas were considered, particularly financial.

2 Background of the forecasting problem

2.1 Theoretical aspects of the forecast

In general, one problem of forecasting time series consists in predicting future values for a time series based on past information. Formally, the forecasting can be defined as the search for or synthesis of a function F, such that (Moro 2002)

$$\vec{x}(t+1) = \vec{F}(\vec{x}(t), \vec{x}(t-1), \ldots, \pi_1, \ldots, \pi_k) \tag{1}$$

where, $\pi_1, \ldots, \pi_k$ are a set of independent parameters of the time variable. Other exogenous time-dependent variables can also be included in equation 1. In practice, the forecasting becomes a problem

of approximating a given function as precisely as possible (Bengio et. al. 1995). To be able to quantify the performance there are various metrics for the forecasting error (Khotanzad & Abaye 1997).

Traditionally, conventional (parametric) statistical models have been used to carry out the forecasting tasks, although Artificial Neuronal Networks (ANN) can be far more adequate for problems having nonlinearities, in addition to their capacity for extrapolating and approximating multivariable data and generalizing from examples (Costa & Ribeiro 1999).

2.2 Application of neuronal networks to forecasting problems

Since the forecasting problem consists in predicting the future value of one or a set of variable based on past patterns, interest must be focused on temporal processing with ANNs. Basically, the ANNs carry out input/output mapping, and to that end they have the ability of representing nonlinear functions arbitrarily (Martín del Brío & Sanz 2001). To carry out the temporal handling of the variables there are two main schemes, one that uses the external handling of the temporal part (delay line) plus a static conventional network (multilayer perceptron (MLP)), and another one that makes use of some properly dynamic network (with internal handling of the temporal variable). Diverse forecasting works are known using MLPs, but they have been limited in temporal processing because they occupy as many input neurons as present and past samples are needed, which is inefficient in terms of network design because of the large number of parameters. On the other hand, dynamic networks, whose output also depends on previous inputs or the system's history, are based on storing a state of the system, either in local internal memories (neuronal) or in the same network recurrence (with feedback connections). Recurrent systems are based on their connection topology which incorporates delays. There are spatially recurrent networks, with connections between neurons, and locally recurrent networks, with internal connections at one neuron. There are various elemental memory mechanisms for temporal neuronal processing (Principe et. al. 2000). The simplest is the "Delay Line PE", which implements in memory a window of length D samples, including the current one. After the limit of samples that the time window can hold (D), the stored values are lost. Other memory mechanisms alter the value of what is stored, but are much more flexible in the temporal aspect, because they can change their effect, modifying parameters without having to alter the topology. For example, there is the "Context PE," which has a feedback connection whose value weights the previous y(n-1) output and produces an exponential memory. Also, there are more advanced memory mechanisms such as the "Gamma Memory PE" (Principe et. al. 2000), which combine the basic memory mechanisms (delays) with an exponential filter type feedback.

Normally, the local dynamic capacities (neuronal) are integrated within a network with more classical topologies, such as the MLP. This is the case of the "time lagged feedforward networks" (TLFN), which are locally recurrent dynamic ANNs that integrate linear filter structures within a feedforward ANN to extend the network's nonlinear capacity with the representation of time. This paper details solutions using TLFN topologies, which are easier to train than spatially recurrent networks and handle the temporal dependence flexibly.

2.3 Forecasting electric load

In the case of the forecasting of electric load, some typologies must be considered, depending on the time interval involved (Gavrilas 2002): long term (intervals of years, systems development planning); medium term (months to one year, maintenance and expansion planning); short term (one day to one week; for a balance between energy load and generation as an efficient way of controlling the system's operation and the electric market); and very short term (for example, ten minutes, for a balance between load and generation during shorter time intervals).

2.4   Methods of analysis of input variables

Carrying out an analysis of input variables consists in studying the contribution of each variable to the result of the forecasting model. This is always useful in practice because it makes it possible to eliminate inputs with little information to describe the output or redundancy between the variables (Bishop, 1995). The steps to be followed are:

- Determine all the candidate variables: based on what an expert in the area of the problem to be solved indicates or on what is stated in the literature.

- Determine linear correlations and statistical analysis: relations between the candidate variables and the output variable, and between themselves, and the behavior of the variable to be forecast is visualized. If there is no linear correlation between two variables, it does not mean that that there are no nonlinear correlations between them (Spiegel, 1997).

Following that, what is mentioned by (Refenes et. al. 1997) can be used, which consists in:

- Building a model that captures the influence of time on the variable to be forecast and the most significant linear dependencies on exogenous variables: it implies identifying the temporal structure using univariate analysis of time series (Hagan & Behr 1997) (García & García 2002) and identifying exogenous influences using multiple linear regression.

- Capturing residual nonlinear dependencies: it is verified whether added value is produced by including additional variables which, even though they may have been rejected by previous linearity criteria, have nonlinear influence. It uses iterative search strategies (forward, backward) and selection criteria, using ANN as a tool.

2.5   Conceptual aspects of preprocessing

Although in theory all ANNs have arbitrary mapping capacities between sets of variables, it is convenient to normalize the data before carrying out the training, to compensate for the inevitable scaling and variability differences between the variables. The above can have a significant impact on the system's performance, so it has to be considered (Bishop 1995).

3   Neuronal modelling of the problem

3.1   Selection of candidate variables

The a priori knowledge of the models determined the following candidate variables: historical electric loads, meteorological (temperature), macroeconomic (GDP, CPI, HSI (hourly salary index)) (Torche 1998) (Gavrilas 2002), demographic (population, housing) (Instituto Nacional de Estadística (INE) 2002b), and referential (month of the year, kind of month) (Karady 2001).

3.2   Selection of input variables

The first phase of the selection method was applied to the candidate variables of the four companies by means of a linear model. An initial analysis was carried out (linear correlation and basic statistics to inspect the series). To determine the past load samples having the greatest influence on the present sample, autocorrelation and partial autocorrelation were used Based on the model obtained, exogenous variables were added (with a forward search strategy). Multiple linear regression was used, measuring the performance with the corrected determination coefficient, ($\overline{R}^2$) (Pérez 2001), which the bigger it is, the

better the quality of the regression. The exogenous variable that yields the largest $\overline{R}^2$ in the corresponding iteration is chosen, provided the improvement achieved is equal to or better than 5% (improvement and value added criterion). The linear models obtained for each company are (with linear function g):

- Company A:
$$C(t) = g(C(t-1), C(t-2), C(t-12), HSI(t)) \tag{2}$$

- Company B:
$$C(t) = g(C(t-1), C(t-2), Popul(t)) \tag{3}$$

- Company C:
$$C(t) = g(C(t-1), C(t-13), Popul(t), Temp(t) \tag{4}$$

- Company C:
$$C(t) = g(C(t-1), C(t-2), C(t-10), Hous(t)) \tag{5}$$

Where C (Load) corresponds to the electric load, and the other variables were defined in section 3.1. As can be seen, for the candidate variables to generate a single "universal and portable" model for the four companies that are studied, they have variable incidence in each case. A general model is developed that has as inputs the past loads and the exogenous variables having the largest linear impact.

- The "CPI" variable was excluded because it does not appear in the linear models. In spite of not appearing in the linear models, the "GDP" variable was not excluded because of a priori knowledge. The "Temperature" variable was excluded because even though it appears in one of the linear models, in general it has a low linear correlation with the load.

- For the input values, gamma memory was used in the input layer to reflect delay.

- In view of the delay with which BT1 customers react to changes in the macroeconomic and demographic variables, the Initial Neuronal Model assumes that the present load has at least one month of delay with respect to the exogenous variables.

3.3   Definition of training and test sets

The following sets of data are defined:

- Learning Set. Inputs: historical load data and exogenous variables (January 1999 through November 2001). Outputs: historical load data from February 1999 through December 2001.

- Cross-Validation Set. Inputs: historical load data and exogenous variables (December 2001 through May 2002). Outputs: historical load data from January 2002 through June 2002.

- Test Set. Inputs: historical load data and exogenous variables (June 2002 through November 2002). Outputs: historical load data from July 2002 through December 2002.

3.4   Data preprocessing

This preprocessing consisted in depurating the available information both for training and for testing the ANN. It was applied to the defined candidate variables (section 3.1).

3.5 Construction of neuronal network for electric load forecasting by low voltage clients

Based on the selection of variables made, the network was modeled with short term memory mechanisms. A training and test were made separately for each company, but applying a general model to the four companies. It consisted in a focalized TLFN (FTLFN) with gamma memory mechanisms in the input layer, in view of its versatility. Since gamma is a unifying system of the two basic memory mechanisms (delay line and context PE), the use of their respective potential is not discarded a priori, since both can be seen as special cases of gamma memories. A hiden layer was used. It will be trained with Backpropagation (static BP with optimization method of descent along the gradient) (Freeman & Skapura 1993) (Martín del Brío & Sanz 2001), since it is a locally recurrent and focalized topology, and also because the μ values (for the gamma memory) will be fixed in an arbitrary but convenient manner, i.e. they will not be adapted, using the criterion of (Principe et al. 2000). The training parametrs are: a momentum of value 0.7 and a "path size" of value 0.1 for the weights that connect the input layer with the hidden layer, and 0.01 for the weights that connect the hidden layer with the output layer (chosen on the basis of the initial tests). In the input layer there will be as many input neurons as there are input variables. The hidden layer is nonlinear, with a "hyperbolic tangent" activation function. The number of hidden neurons will be adjusted based on successive tests to optimize the model. The output layer was chosen to be linear. This layer has one neuron (since there is only one output, the load at instant t) and it will be connected to the hidden layer. The output (still normalized) will be postprocessed. Cross-validation was used to test the quality of learning and avoid overtraining. In this case the training is stopped if the mean square error (MSE) in the cross-validation set does not improve after 100 time-periods (Martín del Brío & Sanz 2001). If the condition is not fulfilled, every component will be submitted to the ANN a maximum of 1000 time-periods. The cross-validation was applied every five training time-periods to measure the corresponding error with a certain stability and avoid sudden oscillation biases. Since the initial weights are random, the results of the training can differ when carried out more than once. For that reason, the network will be trained five times, saving the best weight combination (minimum MSE). To assess the model's viability, a final evaluation is made with the test set. The forecast horizon will be one month (forecasting the following month noniteratively, so as not to carry over probable forecasting errors that may distort the results). Thus, for the training set, a forecast will be made for thirty five months. For the test set the forecast will be made for six months, all this with the forecast horizon mentioned earlier. The forecast is always for the following month. For the training set with BP, the MSE will be calculated successively, defined as follows:

$$MSE = (\sum_{j=0}^{P} \sum_{i=0}^{N}(desired_{ij} - forecast_{ij})^2)/(N*P) \qquad (6)$$

where P is the number of output neurons; N is the number of data set components; $forecast_{ij}$ is component i output in output neuron j; $desired_{ij}$ is desired component i output in output neuron j.

Together with a good approximation, it is desirable for the ANN to give as a result an adequate correlation (absolute value between 0.5 and 1.0), thus reflecting the similarity in the shapes of the desired and forecast output curves. In this way, MSE, MAPE, MAE and correlation (r) will be measured for the cross-validation and test sets (Khotanzad & Abaye 1997). The single Initial Neuronal Model proposed for the four companies is (with nonlinear h function (includes operations performed by gamma memory and ANN on inputs)):

$$C(t) = h(C(t-1), GDP(t-1), HSI(t-1), Popul(t-1), Hous(t-1)) \qquad (7)$$

In the initial model, greater memory depth M (greater number of samples) will be given to past loads and to the exogenous variable of highest linear impact (linear correlation) on the load. For the four companies, this variable was the population. Based on this and on the tests made, the values of for the gamma memories of the corresponding input neurons were fixed as shown in Table 1. After testing

the performance of this model, adding the rest of the candidate variables to the network is evaluated, considering that nonlinear relations with the load could still be captured.

Variable	μ
Historical Load	0.8
HSI	1
GDP	1
Population	0.8
Housing	1

Table 1: Values of μ for the gamma memories of the Initial Neuronal Model

As definition of the model's requirements we have: cross-validation MSE and test less than 0.1; training MSE less than that of cross-validation; for the cross-validation and test the correlation must be greater than 0.5; for cross-validation and test MAPE must be less than 5%. The number of hidden layer neurons was adjusted making successive tests, getting an optimum number of five neurons.

The initial neuronal model is illustrated in figure 1. When three output taps were used in the gamma memories, for the "resolution vs. memory depth" compromise, the optimum values shown in table 2 were obtained.
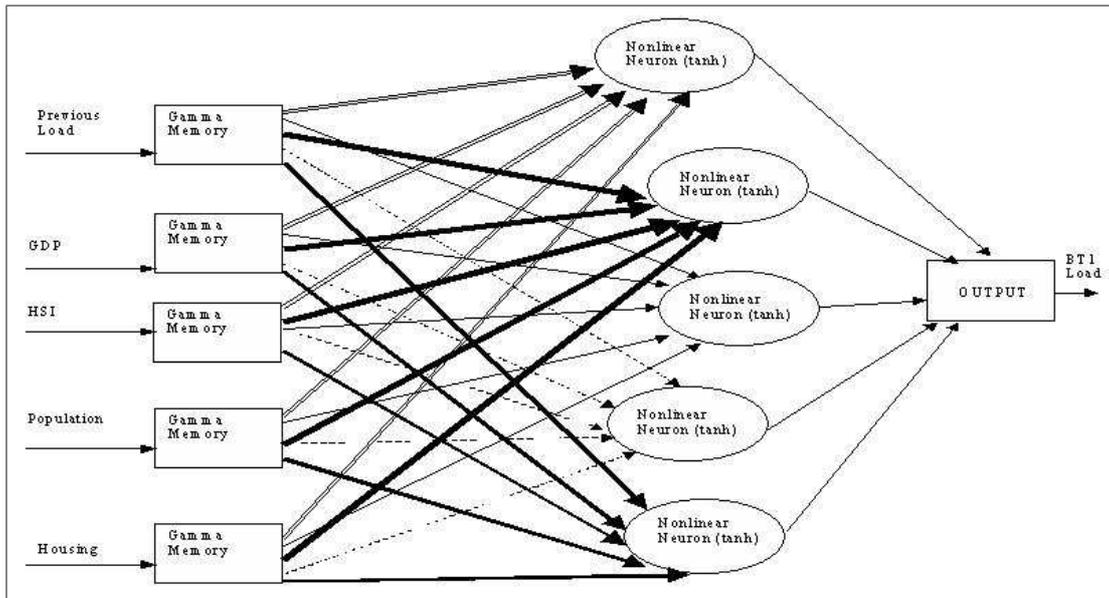

Figure 1: Initial Neuronal Model

The results were satisfactory. Also, the model (single) was portable in topology and μ values (fixed) for the four companies. The only difference is that of the final synaptic weights.

To depurate the initial neuronal model and get residual nonlinear relations of load with other candidate input variables not included so far, the following alternatives were tested: initial model adding "CPI;" initial model adding "Temperature;" initial model adding referential variable "Month of the Year;" initial model adding referential variable "Kind of Month;" and alteration of initial model using as basis the current samples of the exogenous variables. In each case, different combinations of μ were tried, but the results showed no improvement. This depuration attempt was made, in the first place, with a forward search strategy (see section 2.4). No improvement was obtained when either the backward strategy (Refenes et. al. 1997) was used or when variables were removed from the initial model. Therefore, the Final Neuronal Model is the same as the Initial Neuronal Model of equation 7.

Variable	Resolution (R $=\mu$)	Taps (D)	Depth ($M = D/R$)
Historical Load	0.8	3	3.75
HSI	1	3	3
GDP	1	3	3
Population	0.8	3	3.75
Housing	1	3	3

Table 2: Resolution vs. memory depth compromise Initial Neuronal Model

3.6 Multiple linear regression model for the case studied

This model makes an estimation of the dependent variable Y, given the values for a set of independent variables x_i. The general model is given by (Pérez 2001) (Spiegel 1969):

$$Y = b_0 + b_1 * x_1 + b_2 * x_2 + \ldots + b_k * x_k + u \tag{8}$$

where the coefficients b_i represent the effect of the explanatory variables x_k on the independent variable; b_0 is a constant (or independent) term of the model, and u is the residue or error term of the model. When T observations are available in time, equation 9 can be written as follows:

$$Y = b_0 + b_1 * x_{1t} + b_2 * x_{2t} + \ldots + b_k * x_{kt} + u_t \tag{9}$$

where, $t = 1, 2, 3, \ldots, T$.

A multiple linear regression model was constructed using as input the same variables of equation 7. That is, for this case the regression model is defined formally with a dependent variable: $load(t)$; independent variables: $load(t-1)$, $GDP(t-1)$, $HSI(t-1)$, $population(t-1)$, $housing(t-1)$. When the regression is carried out, the coefficients of table 3 are obtained.

Variable	Company A	Company B	Company C	Company D
Constant b	-105058393.5	-27666403.6	-29892228	-28500210.8
Hist. Load	-9.02E-02	-0.119	7.65E-02	-0.19651476
GDP	0.803	-0.18	2.036	-1.29619841
HSI	-97417.432	207055.599	-97524.597	83108.4594
Population	1068.58	0	0	0
Housing	-1718.429	211.062	447.505	455.258917

Table 3: Regression Coefficients by Company

4 Comparative analysis with linear regression model

The results obtained with the neuronal model are compared below with those of the multiple linear regression model already defined. The mean absolute percent error (MAPE) of the sets representative of the ANN model and the regression model were compared. In this way, for each company:

- The MAPE of the regression model was calculated based on all the existing samples so as not to bias (avoid harming or favoring the results of this model) the results in particular samples.

- The MAPE of the neuronal model was calculated for the samples of the cross-validation and for the samples of the test set.

- The MAPE of the neuronal model does not consider the training set, because the network, due to the learning reinforcement itself and the constant error minimization, will produce an almost exact forecast, introducing considerable bias in the results and distorting the comparison.

Table 4 shows a comparison of the results by company:

	Neuronal Cross-Validation	Neuronal Test	Regression
Company A	2.71	2.54	2.77
Company B	1.74	3.03	3.17
Company C	6.95	3.08	3.69
Company D	3.18	5.38	2.71

Table 4: Comparison of Results (MAPE) for the Companies Studied

The previous comparison and its results show that the neuronal model developed was in general superior in accuracy to the regression model, except in specific cases in which the ANN had slight forecasting problems (Company C cross-validation and Company D test). The cross-validation of Company D shows a larger error, but with a slight difference compared to the regression. In the remaining cases the ANN approximation was better than that of the regression model. It may be concluded that the ANN, by using nonlinear relations that affect the electric load, can decrease the forecasting error, giving a better approximation of the variable that is being studied than the statistical model used.

5   Conclusions

This paper and its results show that the ANN represent a powerful tool for decision making in electric distribution companies. This is based on a methodological selection of variables, a prior study of the problem to be solved, and the processing that the ANN can make in the temporal aspect. In this respect, it was decided to construct the neuronal model based on a typology that potentiates the nonlinear static mapping capacity of conventional neuronal topologies (such as the multilayer perceptron), with the handling of time series by means of the "Time Lagged Feedforward Network" (TLFN). To implement it, a versatile memory mechanism (gamma memory) was used, which allows minor adjustments to be made without making changes in the network's general structure, and which generalizes the basic short term memory structures. As future development, forecasting the load of high voltage (AT) customers can be attempted. This is more sensitive to price changes and is more exposed to unexpected functional changes (for example, industry breakdowns) and changes in the macroeconomic variables. The results obtained fulfilled the main and specific objectives of the work, which implied having a reliable ANN model for predicting the monthly electric load of BT1 customers. Also, the ANN model had better accuracy than the classical statistical model used. All this means that the model constructed makes it possible to determine the medium term load with acceptable accuracy, as required by the electric distribution companies for the system's planning and expansion. The gamma networks are more versatile for modelling nonstationary time series. Handling of the feedback parameter gives the developer an additional element for adjusting the network, disregarding the increase or decrease of the neurons of the hidden layers, which often is ambiguous. Support by linear statistical techniques and time series is an important tool for selecting variables, making it possible to detect linear dependencies prior to the construction of the ANN, leading to focusing the test/error of such network on finding nonlinear dependencies, with a large part of the linear modelling already solved. The electric customer reacts with delay or inertia to changes in the macroeconomic variables, and the load reacts with inertia to changes in the demographic variables. It was shown that the present electric load has a delay of at least one month in its "reaction" to exogenous variables. From the above it is seen that if short term memory mechanisms were used for the neuronal

modelling, the delays that must be considered should be handled carefully. Otherwise the training would not be useful.

References

[1] Bengio, S., Fesant, F., Collobert D., *A Connectionist System for Medium - Term Horizon Time Series Prediction*, International Workshop on Applications of Neural Networks to Telecommunications, IWANNT, Stockholm, Sweden, 1995.

[2] Bishop, C., *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[3] Costa, N., Ribeiro, B., *A neural prediction model for monitoring and fault diagnosis of a plastic injection moulding process*, Centro de Informática e Sistemas da Universidade de Coimbra, 1999.

[4] Freeman, J. and Skapura D., *Redes Neuronales: Algoritmos, Aplicaciones y Técnicas de Programación*, Versión en español de García-Bermejo, R. Joyanes, L. Addison Wesley / Díaz de Santos, Wilmington, Delaware, EUA, 1993.

[5] García, J. and García, F., *Econometría. Tema 5. Autocorrelación*, Curso 2002/2003. Universidad de Huelva, 2002.

[6] Gavrilas, M., *Neural Network based Forecasting for Electricity Markets*, Technical University of Iasi, Romania, 2002.

[7] Hagan, M. and Behr, S., "The Time Series approach to short term load forecasting", *IEEE Transactions on Power Systems*, Vol. PWRS-2, NO. 3, 1997.

[8] Instituto Nacional de Estadísticas (INE 2002b), *Censo 2002. Síntesis de Resultados*.

[9] Karady, G., *Short - Term Load Forecasting using Neural Networks and Fuzzy Logic*, Arizona State University. Power Zone, 2001.

[10] Khotanzad, A., Abaye A., "ANNSTLF - A Neural Network Based Electric Load Forecasting System". *IEEE Trans. on Neural Networks*, Vol. 8, No. 4, 1997.

[11] Martín del Brío, B., Sanz Molina A., *Redes Neuronales y Sistemas Borrosos*, 2da. Edición, Ra-Ma, Madrid, España, 2001.

[12] Moro, Q., *Series Temporales y Redes Neuronales Artificiales*, Departamento de Informática Universidad de Valladolid. Pérez, C. (2001), *Técnicas Estadísticas con SPSS*, Prentice Hall, Madrid, España, 2002.

[13] Principe J., Euliano N., Lefebvre W., *Neural and Adaptive Systems: Fundamentals Through Simulations*, John Wiley and Sons, New York, 2000.

[14] Refenes, A., Burgess, A., Bentz, Y., "Neural Networks in Financial Engineering: A Study in Methodology". *IEEE Transactions in Neural Networks*, Vol 8; No. 6, 1997.

[15] Spiegel, M., *Estadística*, Serie de Compendios Schaum. Libros McGraw-Hill, Colombia, 1969.

[16] Torche, A., *Contabilidad Nacional. NúmerosIndices. Desestacionalización y Trimestralización*, Trabajo Estadístico Número 63. Pontificia Universidad Católica de Chile. Insituto de Economía. Santiago, 1998.

Danilo Bassi and Oscar Olivares
Universidad de Santiago de Chile
Departamento de Ingeniería Informática
Santiago, CHILE
E-mail: dbassi@ieee.org, oolivare@yahoo.com