Journal of Modern Applied Statistical Methods

Volume 5 | Issue 2 Article 24

11-1-2005

Statistical Methods and Artificial Neural Networks

Mammadagha Mammadov
Anadolu University, mmammadov@anadolu.edu.tr

Berna Yazici
Anadolu University, Turkey, bbaloglu@anadolu.edu.tr

Şenay Yolaçan
Anadolu University, senayyolacan@anadolu.edu.tr

Atilla Aslanargun
Anadolu University, aaslanar@anadolu.edu.tr

Ali Fuat Yüzer
Anadolu University, afyuzer@anadolu.edu.tr

See next page for additional authors

Statistical Methods and Artificial Neural Networks

Authors
Mammadagha Mammadov, Berna Yazici, Şenay Yolaçan, Atilla Aslanargun, Ali Fuat Yüzer, and Embiya Ağaoğlu

Journal of Modern Applied Statistical Methods
November, 2006, Vol. 5, No. 2, 495-512

1538 – 9472/06/$95.00

Statistical Methods and Artificial Neural Networks

Mammadagha Mammadov Berna Yazıcı Şenay Yolaçan
Atilla Aslanargun Ali Fuat Yüzer Embiya Ağaoğlu
Anadolu University

Artificial Neural Networks and statistical methods are applied on real data sets for forecasting, classification, and clustering problems. Hybrid models for two components are examined on different data sets; tourist arrival forecasting to Turkey, macro-economic problem on rescheduling of the countries' international debts, and grouping twenty-five European Union member and four candidate countries according to macro-economic indicators.

Key words: Time series, ARIMA, neural networks, hybrid models, logistic and probit regression, rescheduling and non-rescheduling of the international debts, Kohonen nets, cluster analysis, Maastricht criteria

Introduction

Artificial Neural Networks (ANN) is efficiently being used as an alternative of statistical methods for different problems like estimation, classification, clustering analysis, sample recognition and etc. Since ANN models are usually nonlinear, those models give better estimates in application.

In this study for real data sets forecasting, classification and cluster problems and hybrid models for two components are examined. The statistical models (ARIMA) used for time series analysis are usually linear. Therefore, using ANN models that can impress the nonlinear structure has a great importance. Zhang et al. (2001) presented a recent review in this area. There are numerous studies to compare the performances of ANN and traditional time series techniques. For example, the empirical results in Ansuj et al. (1996), Caire et al. (1992), Chin and Arthur (1996), Hill and O'Connor (1996), Kohzadi et al. (1996), Maier and Dandy (1996) showed that the ANN gave improved results in terms of forecasting accuracy.

Recently, an approach was recommended that uses ARIMA and ANN models together for time series forecasts (Tseng et. al., 2002; Zhang, 2003). Ginzburg and Horn (1994) and Pelikan et. al. (1992) proposed to combine several feedforward neural networks to improve time series forecasting accuracy. Wedding and Cios (1996) described a combining methodology using radial basis function networks and Box and Jenkins models. Voort et al. (1996) used a Kohonen self-organizing map as an initial classifier; with each class having an individually tuned ARIMA model associated with it. Luxhoj et al. (1996) presented a hybrid econometric and ANN approach for sales forecasting. Wang and Leu (1996) developed neural network trained by features extracted from ARIMA analysis. Their results showed that the neural network trained by different data produced better predictions than otherwise trained by raw data. Results from Su et al. (1997) showed that the hybrid model

Mammadagha Mammadov is Professor of Statistics. Email: mmammadov@anadolu.edu.tr Berna Yazici is Assistant Professor of Statistics. Email: bbaloglu@anadolu.edu.tr. Senay Yolacan is Research Assistant. Email: senayyolacan@anadolu.edu.tr. Atilla Aslanargun is Assistant Professor of Statisitcs. Email: aaslanar@anadolu.edu.tr. Ali Fuat Yüzer is Professor of Statistics. Email: afyuzer@anadolu.edu.tr. Embiya Agaoglu is Professor Statistics. Email: eagaoglu@anadolu.edu.tr

produced better forecasts than either the ARIMA model or the neural network by itself. Thus, the combined model has produced promising results in these studies.

According to these results on hybrid models, constructed by combining ANN and statistical models or different ANN models makes more accurate time series forecasts. Choosing the hybrid model depends on the statistical and mathematical structure of the data. Of course, usually this relationship can not be explained. Therefore, the components of the hybrid models are chosen by experiments.

Economics, finance, and business have great importance among the problems examined using ANN techniques (Wilson & Sharda, 1992, Kuan & White, 1994, Sharda, 1994, Wong et al., 1994). There are an increasing number of articles on economics that applied ANN models in recent years. Leung et al. (2000) compared the performance of different forecasting techniques and general regression in forecasting exchange rates. Kim et al. (2004) mentioned the usefulness of neural networks for early warning system of economic crisis. They studied the Korean economy as an example of world economic crisis.

One of the other macroeconomic problems that are examined using ANN is country's international debt problems. There are also numerous researches using statistical methods on this subject. Frank and Cline (1971) used multiple discriminant analysis to predict debt-servicing difficulties. Dhonte (1975) used principal components analysis to obtain a description of a country's debt position. Feder and Just (1977) used logit analysis in determining debt servicing capacity. Yoon-Dae Euh (1979) made multiple regression analysis in determining of a country's creditworthiness. Kharas (1984) made probit analysis to assess the probability of a country's becoming uncreditworthy. Cooper (1985) used canonical correlation analysis to examine the relationship between country creditworthiness and a country's recent economic performance.

Rescheduling and non-rescheduling of the international debts of countries is examined. As a result of classification, more successful architectures and algorithms are defined.

Another macroeconomic problem taken into account is whether the Economic and Monetary Union (EMU) appears to be a wholly homogenous group of countries due to Optimal Currency Area (OCA) or not. This is an important question because the sustainability of EMU depends on the existence of a reasonable degree of homogeneity (Artis & Zhang, 2001). A long debate on this problem has taken place after the first publication on optimum currency area among different countries by Mundell (1961) and followed by McKinnon (1963), with important elaborations by, among others, Kenen (1969) and Krugman (1990). Artis and Zhang (2001) examined the status of the EMU member countries by the time the euro was launched with criteria being measured according to optimal currency area criteria using a cluster analysis approach.

In this article, homogeneities in the actual and candidate members of the EMU are examined using Neural Network techniques and cluster analysis, traditional statistical technique.

Artificial Neural Network Algorithms

In this part of the study, a brief discussion is given for ANN as an alternative of statistical methods for estimation, classification and cluster analysis.

Standard Backpropogation (BPM)

First, multi-layer feedforward perceptron (MLP) ANN models that are widely used for estimation and classification problems are taken into account. Backpropagation is the widespread approximation approach for training of the multi-layer feedforward neural networks. The main idea here is minimizing the sum of square error at each epoch (Bishop, 1995; Haykin, 1999). In fact, this algorithm is an application of gradient descent methods at numerical optimization to error function of which the variables are the weights (Nocedal & Wright, 1999). At each epoch the weight vector is changed in the direction of the error function gradient in defined ratio at the currency point and the value of error is lessened:

$$w_{k+1} = w_k - \alpha \cdot g_k$$

where w_k is the current vector of weights and biases, g_k is the current gradient of error function at the point w_k, and α is the learning (training) rate.

In the simple gradient method there are some difficulties about local minimum and speed of convergence. Thus, the momentum term is added to weight change formula:

$$\Delta w_{k+1} = -\alpha \cdot g_k + \mu \Delta w_k , \qquad (1)$$

where $\Delta w_k = w_k - w_{k-1}$ is the weight change in the previous iteration, μ is the momentum coefficient. The spread algorithm with momentum is called the standard backpropagation algorithm. In this algorithm, the better choice of α and μ constants speed of convergence and stability of the algorithm (Yu & Chen, 1997; Qian, 1999; Bhaya et al., 2004).

Resilient Backpropagation (RP)

MLP's typically use S-shaped sigmoid activation functions in the hidden layers. For this reason, their slope must approach zero as the input gets large. The purpose of the resilient backpropagation (RP) training algorithm is to eliminate these injury effects of the magnitudes of the partial derivatives. In RP algorithm the magnitude of the derivative has no effect on the weight update. Only the sign of the derivative is used to determine the direction of the weight update. If the derivative is zero, then the update value remains the same. Whenever the weights are oscillating the weight change will be reduced. If the weight continues to change in the same direction for several iterations, then the magnitude of the weight change will be increased.

There are some different versions of backpropagation ANN algorithm like Conjugate Gradients and Newton method. Those algorithms which are used for the experimental studies in the article are briefly given in this section.

Conjugate Gradients (CG) Algorithm

This is a special version of conjugate direction method. For a starting point $w_0 \in R^n$, the method defined by the formulas below called conjugate direction method (Nocedal & Wright, 1999):

$$w_{k+1} = w_k + \alpha_k p_k , \quad \alpha_k = -\frac{p_k^T g_k}{p_k^T H p_k} \qquad (2)$$

Here, α_k is defined by the minimum problem of one variable function of $\varphi(\alpha) = F(w_k + \alpha p_k)$ and $\{p_i\}_{i=1}^{n}$ conjugate directions. For any starting point $w_0 \in R^n$, the sequence $\{w_k\}$ generated by the conjugate direction algorithm converges to the minimum point w^* of the problem (2.3) in at most n steps (Nocedal & Wright, 1999). The conjugate gradient method is a conjugate direction method; start out by searching in the gradient direction on the first iteration

$$p_0 = -g_0 \qquad (3)$$

The conjugate current p_k vector is calculated by using the previous p_{k-1} vector and current gradient

$$p_k = -g_k + \beta_k p_{k-1} \qquad (4)$$

Coefficient β_k is selected by the condition of p_{k-1} and p_k vectors' being conjugate with respect to the symmetric positive definite $n \times n$ Hessian matrix H ($p_{k-1} H p_k = 0$):

$$\beta_k = \frac{g_k^T H p_{k-1}}{p_{k-1}^T H p_{k-1}} . \qquad (5)$$

There are various CG algorithms depending on different calculation formulas of β_k coefficients, for example Fletcher–Reeves, Polak and Ribere conjugate gradient algorithms. Some important examples are due to Bishop (1995), Haykin (1999), and Nocedal and Wright (1999).

Another algorithm of CG algorithms is Scaled Conjugate Gradients (SCG) Algorithm. CG algorithms reviewed above makes a line search at each iteration and makes calculations very expensive. Moller (1993) put forward scaled conjugate gradient algorithm that does not use line search procedure in traditional CG algorithms. The idea of SCG is to combine the model trust region approach with the conjugate gradient approach (Bishop, 1995; Nocedal & Wright, 1999).

Quasi-Newton (QN) Algorithms

Newton's method is

$$w_{k+1} = w_k - H_k^{-1} g_k , \tag{6}$$

where H_k is the Hessian matrix in current point w_k. Newton's method often converges faster than conjugate gradient methods. Unfortunately, it is complex and expensive to compute the Hessian matrix. Quasi-Newton (or secant) methods are based on Newton's method, but don't require the calculation of second derivatives. They update an approximate Hessian matrix at each iteration of the algorithm (Nocedal & Wright, 1999). (The update is computed as a function of the gradient.)

The Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm is one of the successful QN algorithms. The one step secant (OSS) algorithm is an attempt to bridge the gap between the CG algorithms and Quasi-Newton algorithms. This algorithm accepts that the previous Hessian is the identity matrix and so inverse matrixes are not calculated to choose the new search direction. This algorithm requires very light storing and requires less storing at each step than the CG algorithms.

The Levenberg-Marquart (LM) algorithm is one of the required QN algorithms (Bishop, 1995). Hessian matrixes are not calculated and not taken into account of its being second order, the Hessian matrix can be approximated as $H = J^T J$, where J is the Jacobean matrix that contains first derivatives of the network errors with respect to the weights and biases. The gradient can be computed as $g = J^T \varepsilon$, where ε is a vector of network error. The Levenberg-Marquart algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$w_{k+1} = w_k + [J^T J + \mu I]^{-1} J^T \varepsilon . \tag{7}$$

This algorithm runs fast for moderate-dimensioned feedforward neural networks for regression problems.

Radial Basis Function Networks (RBFN)

RBFN (Haykin, 1999; Bishop, 1995) are also used, aside from MLP networks, in regression and classification problems. In RBFN, one hidden layer with required number of units is enough in order to model a function. The activations of hidden (radial) units are defined depending on the distance of the input vector and the center vector. Typically, the radial layer has exponential activation functions and the output layer a linear activation function.

Education is made in three stages in RBFN. In the first stage by unsupervised education, radial basis function centers are optimized using all $\left\{x^{(i)}\right\}$, $i = 1, 2, ..., N$ education data. Centers can be assigned by a number of algorithms: Sub-Sampling, K-means, Kohonen training, or Learned Vector Quantization. In the second stage σ_j, $j = 1, 2, ..., p$ deviation (or width) parameters can be assigned by algorithms such as Explicit, Isotropic, K-nearest neighbor. In the third stage of education, the basis functions that are obtained for adjusting the appropriate weights for output units are taken fixed and deviation parameters are added to linear sum. Optimum weights are obtained by minimization of the sum of square errors. The output layer is usually optimized using the Pseudo-Inverse technique.

MLP with a defined architecture is given by the appropriate weights and the biases of the units, but in RBFN, it is given by the center and the deviation of the radial units and by the weights and biases of the output units. Since point is given by n coordinates in n-dimensional space, the number of the coordinates are equal to the linear input units n. So in ST Neural Network software, the coordinates of the center radial unit are taken as weights, and the deviation of the radial unit is taken as bias. As a result, radial weights denotes the center point, radial bias denotes the deviation.

Having only one hidden layer and making faster education than MLP can be taken as advantages of RBFN. The linear modeling methods are more useful in output layers of RBFN that removes the difficulties which occur about the local minimums in MLP.

However, RBFN has some disadvantages compared with MLP. In order to correctly modeling a typical function in RBFN, many more hidden (radial) units can be required than appropriate MLP model. That may cause running the model slower and more memory can be required. RBFN is very sensitive to increment of the network dimension and some difficulties can be occurred as a result of increment in the number of input units.

RBFN is unsuccessful in extrapolation in its nature. MLP networks are more successful in extrapolation problems than RBFN. Because, when the input data are far from the radial centers, than the output signal is 0, and this may not show the required result.

Kohonen Networks (Kohonen Self-Organizing Maps (SOM))

The algorithms given above belong to the supervised training rule. That training is evaluated due to a given target. Nevertheless Kohonen networks are fulfilled a topologic structure among the cluster units by unsupervised training. That property is observed as one characteristic of the brain and does not occur at other ANNs. These Networks are based on competitive learning. The neurons of competitive Networks learn to recognize groups of similar input vectors. An output neuron that wins the competition is called a winning neuron. The weights of wining neuron with index j and its specified ith unit neighborhood in training algorithm are updated as follows:

$$w_{iJ}(k+1) = w_{iJ}(k) + \alpha[x_i - w_{iJ}(k)] \qquad (8)$$

where α is the learning ratio. At each epoch of algorithm, the radius R and learning ratio α are changed by decreasing. There are some alternative structures in order to decrease R and α. Kohonen (1989) showed that α is satisfactory for linearly decreasing functions, and geometrically decreasing can produce similar results.

An Application Study on Time Series Using ARIMA, ANN and Hybrid Models

The statistical models used for time series analysis are usually linear. Thus, using ANN models that can impress the nonlinear structure is of great importance. Zhang et al. (2001) presented a recent review in this area. Experimental results with real data sets indicate that a hybrid methodology that combines both ARIMA and ANN models can be an effective way to improve forecasting accuracy achieved by either of the models used separately. G. P. Zhang (2003) explained the reasons for using hybrid models in detail.

According to hybrid model results, combining ANN and statistical models or different ANN models produces more accurate forecasting for time series. Choosing the hybrid model depends on the statistical and mathematical structure of the data. Of course, this relationship can not usually be explained. As a result, the components of the hybrid models are chosen by experiment.

Some articles on tourism forecasting problems mention ANN as having better performance than statistical techniques. Law and Au (1999) found that using the feedforward neural network model to forecast annual Japanese arrivals for travel to Hong Kong outperformed multiple regression models, naive, moving average, and exponential smoothing in terms of forecasting accuracy. The empirical results of Law (2000) showed that BP neural network outperformed regression models, time series models and feedforward neural networks in terms of forecasting accuracy. In his article, Vincent Cho (2003) investigated the applications of three time series forecasting techniques, namely exponential smoothing, univariate ARIMA, and Elman’s model of ANN. Neural networks seem to be the best method for tourist arrival forecasting. Kim et al. (2003) provided a short introduction to the use of Kohonen’s SOM algorithm in tourism and presents a descriptive analysis of the ANN methodology. It provides a research technique that assesses the weighting of different attributes, and uses an unsupervised

ANN model to describe a consumer product relationship.

The ANN approach and the hybrid methodology to time series modeling

For one hidden layer network architecture $n{:}p{:}1$ (n: number of inputs, *p*: number of hidden units and 1: number of outputs), inputs are the observed values of n^{th} previous time points and outputs (targets) are $(n+1)^{th}$ the observed value. ANN are nonlinear functions of previous observations $(y_{t-1}, y_{t-2}, \ldots, y_{t-n})$ to y_t future observations (Zhang, 2003):

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-n}, w) + \varepsilon_t, \qquad (9)$$

where $(y_{t-1}, y_{t-2}, \ldots, y_{t-n})$ are input values, y_t is target value, w are weights of the network, ε_t are the vector of biases at time point t. The $\hat{y}_t$ prediction value is calculated as follows:

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-n}, w) \ . \qquad (10)$$

In training procedure, with the help of different backpropagation algorithms, the parameters (weights and biases) of the network are obtained by getting closer to the minimum value of the sum of the square error $SSE = \sum_{t=n+1}^{N} (y_t - \hat{y}_t)^2$.

In recent articles (Tseng et. al., 2002; Zhang, 2003), more importance is given to hybrid models that are composite of ARIMA and NN models. However, the proposed hybrid model does not always show a better performance. The idea of model combination in forecasting is to use each model's unique feature to capture different patterns in the data (Zhang, 2003).

On time series sampling in this study, ARIMA&NN and different NN&NN hybrid models are evaluated with experimental calculations, with interesting results obtained. As a result of the experiment, it is seen that hybrid models like MLP&MLP and MLP&RBFN with two nonlinear components show better performance for forecasting problems. So, the hybrid model structure of Zhang (2003) can be extended. This means that both of the components of the hybrid model may be nonlinear at the same time (Aslanargun et al., 2007):

$$y_t = y_t^1 + y_t^2 \qquad (11)$$

In this model, y_t is the observation value at time point t, y_t^1 and y_t^2 are linear or nonlinear model components, and superscripts denote the row number of the model. Firstly, the model with 1-indiced is applied to the observation data and $e_t^1 = y_t - \hat{y}_t^1$, then the others are calculated. Here, $\hat{y}_t^1$ is the forecast value of the first model at time point t. If the first model contains m_1 input units, the number of e_t^1 units will be $N - m_1$. If the second model contains m_2 input units, the number of $\hat{y}_t^2$ forecast values will be $N - m_1 - m_2$. In this case, the forecast values appropriate for the second model are calculated as follows:

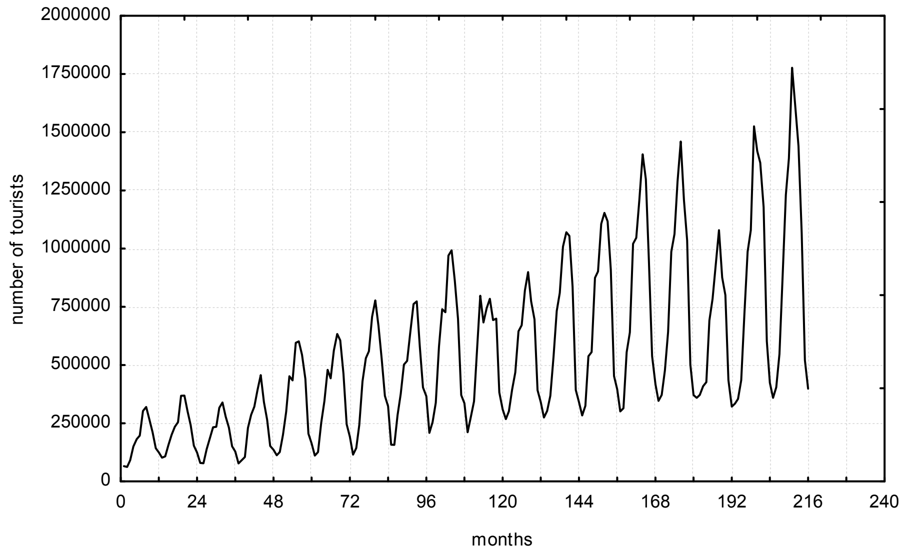

Figure 1. Number of monthly tourist arrivals to Turkey (1984-2001).

$$\hat{y}_t^2 = f_2(e_{t-1}, e_{t-2}, \ldots, e_{t-m_2}) \qquad (12)$$

where f_2 is the function obtained from the second model. The forecast for the combined model is defined as follows:

$$\hat{y}_t = \hat{y}_t^1 + \hat{y}_t^2 \qquad (13)$$

The adjusted forecasts are calculated as the sums of the first model and the second model. The hybrid model with good performance is obtained by the evaluation measure for the forecasting.

Experimental evaluation

In this part of the study, the time series of number of monthly tourist arrivals to Turkey was examined. Appropriate ARIMA, ANN and hybrid models were chosen by doing experiments in order to make forecasts, and these models are also compared. For the analyses STATISTICA Neural Networks and SPSS statistical packages are used.

The data set, the number of monthly tourist arrivals to Turkey between January 1984 and December 2003, is taken from the Republic of Turkey, Prime Ministry State Institute of Statistics (www.die.gov.tr).

The graph of 216 month period for January 1984 to December 2001 is given in Figure 1. Examining Figure 1 it is seen that the series is not stable in variance, after the logarithmic transformation variance stability criterion is provided, and it can also be seen that the series includes seasonal and trend effects. Thus, it was decided that the most appropriate model for this series is ARIMA$(1,1,1)(1,1,0)_{12}$.

The 216 monthly tourist arrivals to Turkey data for January 1984 to December 2001 period was used in training of the network. An evaluation of the model was made dependent on the forecasts for the 24 month period between January 2002 and December 2003. The choice of the best model depends on a comparison of statistics such as the MSE (RMSE) and MAE. Since the initial weight and bias values of the network were random, 150 replications were

Table 1. The performance values for the selected models

Models	MSE	RMSE	MAE
ARIMA(1,1,1)(1,1,0)$_{12}$	3.06E+10	174926.3	137589.4
(12:1)LinearNN	2.93E+10	171046.8	144662.4
(12:1:1)MLP	2.19E+10	147909.3	127838.8
(12:48:1)RBFN	5.40E+10	232280.8	176592.2
ARIMA(1,1,1)(1,1,0)$_{12}$&(6:8:1)MLP	2.79E+10	167068.6	128148.6
ARIMA(1,1,1)(1,1,0)$_{12}$&(6:7:1)RBFN	2.86E+10	169058.0	129721.5
(12:1)LinearNN&(6:8:1)MLP	2.73E+10	165220.0	141377.4
(12:1)LinearNN&(9:15:1)RBFN	2.57E+10	160479.0	139907.8
(12:1:1)MLP&(5:1:1:1)MLP	2.09E+10	144410.7	123104.7
(12:1:1)MLP&(6:9:1)RBFN	2.07E+10	143958.1	123184.7
(12:48:1)RBFN&(12:8:8:1)MLP	4.78E+10	218540.3	160724.3
(12:48:1)RBFN&(12:12:1)RBFN	4.99E+10	223470.6	167694.1

made for the same network structure and the models giving the best forecasts were determined.

Because the tourist arrival data in question included seasonality, after trying many neural networks with different numbers of input units, as expected, the number of input units was determined as 12. Various neural network algorithms with single layer, with one or two hidden layers, multilayer feedforward algorithms and RBFN models were applied on the sample data set. Twelve were lost due to seasonality, so 204 of 216 were used to adjust the weights. In the training stage of the network, data were divided into two parts; 132 of the 204 data were used for training and 72 data were used for validation. This division was used to restrict memorization of the network and provided for better forecasts (Bishop, 1995, Haykin, 1999).

The single layer neural network with (12:1) architecture (12 input units, 1 output unit and 1 bias) showed better performance for the forecast data as a result of adjustment of the suitable weights in training. The neural network that showed the best performance among the MLP for the time series in question was found to be the (12:1:1) MLP. The QN method showed the best performance in 33 epochs. Among the RBF Networks the (12:48:1) RBFN showed the best performance. The hyperbolic-tangent function is applied in the hidden layer and the linear activation function is applied in the output unit. The biases and weights resulting from training are obtained.

Hybrid models such as the ARIMA&MLP, LinearNN&MLP, ARIMA&RBFN, LinearNN&RBFN and MLP&MLP, MLP&RBFN, RBFN&MLP, RBFN&RBFN were taken into account. In order to determine the ARIMA&MLP and the ARIMA&RBFN hybrid models, firstly, the ARIMA(1,1,1)(1,1,0)$_{12}$ model was applied to the data. A series of residuals, the differences between the observed values and the estimates of these models were obtained. In the series of residuals, since a 1st order difference and a 1st order seasonal differences were taken in the ARIMA model, 13 data were lost and 203 data remained. At the next step, the MLP or RBFN models were applied to the series of residuals. The same procedure is applied for other hybrid models.

The hybrid models with the best performance are given in Table 1. The weights and biases are also obtained for all the second components of hybrid models.

Results Part 1

The observed and forecasted values of number of monthly tourist arrivals to Turkey for January, 2002 and December, 2003 period is obtained. Also MSE, RMSE and MAE values for the forecasts are obtained in order to compare the performances of ARIMA, ANN and hybrid models mentioned in previous sections and results are given in Table 1.

As seen in Table 1, due to the MSE and RMSE measures, (12:1:1)MLP&(6:9:1)RBFN,

(12:1:1)MLP&(5:1:1:1)MLP hybrid models and (12:1:1)MLP ANN models respectively; due to the MAE measure (12:1:1)MLP&(5:1:1:1)MLP, (12:1:1)MLP&(6:9:1)RBFN hybrid and (12:1:1)MLP ANN models respectively make the best forecasts. The graph of forecasted values is given in Figure 2.

Among the models with one component (the models that are not hybrid models), due to the MSE and MAE measures, the (12:1:1)MLP showed the best performance, and the (12:48:1)RBFN showed the worst performance. It is known that the RBFN is usually unsuccessful in extrapolation problems (Bishop, 1995). In this manner, the result from Table 1 for the RBFN is an expected result. Although the RBFN makes bad forecasts by itself, as can be seen from Table 1, when it is used as a second component, the (12:1:1)MLP&(6:9:1)RBFN hybrid model showed the best performance when it is used as a second component. When only the RBFN is used, the number of radial units is defined as 48. In other words, when the start data set is used in education and forecast, 48 radial centers are selected. In the (12:1:1)MLP&(6:9:1)RBFN hybrid model, when the second component RBFN is applied to the remaining data, the number of radial centers are decreased to 9. This situation can be explained thus; the differences between the residuals are getting smaller.

An Application Study on a Classification Problem

Interest on macro economic problems using ANN has increased since 1990s. Roy and Cosset (1990) developed a neural network to estimate sovereign credit ratings. Burrell and Folarin (1997) used neural networks in order to maintain competitiveness in a global economy and collected the studies on global economy using neural networks until then. Cooper (1999)

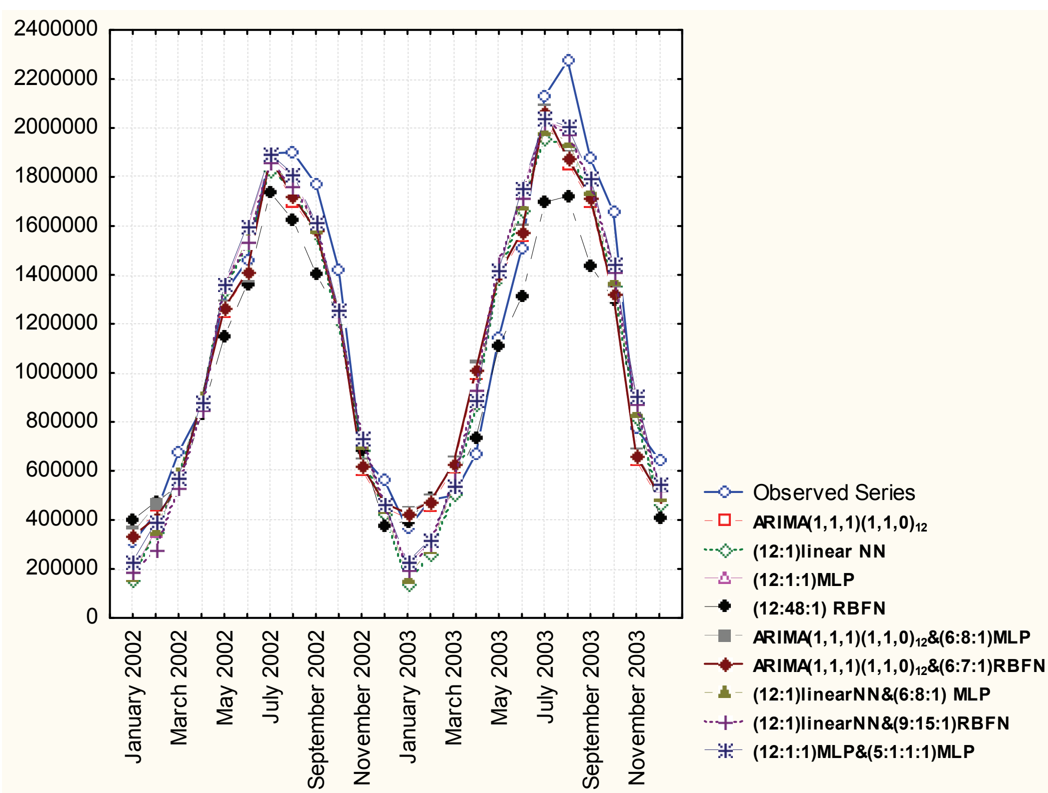

Figure 2. The observed and forecasted values for the January 2002 to December 2003 period.

compared the performance of artificial neural networks with logistic, probit regression and discriminant analysis to classify the countries to two classes; reschedule or nonreschedule the international debts.

Rescheduling and non-rescheduling of the international debts of countries is examined using data for 1982. To analyze the data taken into account, logistic and probit regression among the statistical methods and different backpropogation algorithms of multilayer networks as ANN are used. As a result of classification, more successful architectures and algorithms are defined.

Definition of Macro Economic Problem

Because international debt crisis effects all economical and political balances of countries, in this study, countries' rescheduling and non-rescheduling of their international debts is examined. In 1982, international debt problem became such a big problem that detailed studies required for lending phase of countries (Cooper, 1999). It became an investigation topic whether the countries will pay their debt in time or not, whether they reschedule their debt or not. Hence, it this study countries' rescheduling and non-rescheduling of the international debts is considered as a classification problem and the data for 70 countries of 1982 are used.

The factors that affect the rescheduling and non-rescheduling of the international debts defined as follows and those factors are taken as independent variables for logistic and probit regression, and inputs for ANN:

X_1: Average increase in GDP growth (annual %) over 1960-1982

X_2: The ratio of short term debt (%GDP) to export of goods and services (%GDP)

X_3: Debt service ratio, the interest on total external debt plus amortization on long-term debt as a percentage of the exports of goods and services in 1982.

X_4: The import of cover and equals international reserves divided by total imports in 1982.

The data for this study which were collected from Morgan Guaranty Trust Company and World Bank is taken from Cooper (1999).

The number of input variables is four for MLP model since all those variables are taken as input variables. The output variable can either be 1 for rescheduled debts if probability is equal or greater than 0.5 or 0 for non-rescheduled debts if probability is less than 0.5. So in output layer of the network, there is only one neuron.

Analysis

Cooper (1999) applied logistic, probit, discriminant analysis and gradient descent with momentum with one hidden layer 4:8:1 algorithm for countries divided to 3 groups of 20 units. But, 4:8:1 architectured resilient backpropagation of ANN algorithms that is applied in this study give 100% correct classification.

In this study, logistic and probit regression analysis are made for modeling of the data related with 70 countries. The obtained models are given as follows:

$$L=-3.4610-0.1872X_1+0.0228X_2+0.0869X_3-0.1782X_4$$

$$P=-1.9650-0.0879X_1+0.0117X_2+0.0471X_3-0.0669X_4$$

To the same data, gradient descent with momentum algorithm (GDM), scaled conjugate gradient algorithm (SCG) and resilient backpropagation algorithm (RP) of backpropagation algorithms of ANN are applied. In order to construct models for extrapolation with ANN, it is useful to divide the data to three different groups for training, validation and test. During the process of modeling, training is made by the training data and total square error related to validation data is calculated in each epochs. In a certain epoch, training is stopped when the error related with the validation data start to increase and the weights of the net is determined. If the errors corresponding to training, validation and test data have small values in the same time, the architecture of determined neural net and the algorithm is respected to be good (Demuth & Beale, 1996). The model constructed by this training method, satisfy the opportunity to make more successful extrapolation.

Table 2. Extrapolation for the data consist of 1988-2001 related to Turkey

	Data			
	X_1	X_2	X_3	X_4
1988	2.1	0.8	2.4	2.1
1989	0.03	0.09	2.4	3.5
1990	9.3	1.4	2.3	4.2
1991	0.09	1.3	2.3	3.9
1992	0.06	1.6	2.4	4.4
1993	0.8	2	2.5	4
1994	-5.5	0.8	2.1	4.2
1995	7.2	1.1	0.2	5.7
1996	7	1	1.8	6.4
1997	7.5	0.9	1.8	6.4
1998	3.1	0.9	1.7	7.4
1999	-4.7	1	2.3	9
2000	7.4	1	2	7.4
2001	-7.5	0.4	2.9	6.3

	Extrapolation					
	SCG		GDM		RP	
	(4:8:1)	(4:11:1)	(4:8:1)	(4:11:1)	(4:8:1)	(4:11:1)
1988	0.9931 (1)	0.9990 (1)	0.9945 (1)	0.8473 (1)	0.9899 (1)	0.9937 (1)
1989	0.7857 (1)	0.9017 (1)	0.8438 (1)	0.5536 (1)	0.9694 (1)	0.9806 (1)
1990	0.9981 (1)	0.9994 (1)	0.9982 (1)	0.8648 (1)	0.9940 (1)	0.9995 (1)
1991	0.8176 (1)	0.9764 (1)	0.8912 (1)	0.5742 (1)	0.9535 (1)	0.9053 (1)
1992	0.8255 (1)	0.9828 (1)	0.8981 (1)	0.5862 (1)	0.9516 (1)	0.9110 (1)
1993	0.8991 (1)	0.9969 (1)	0.9704 (1)	0.7703 (1)	0.9819 (1)	0.9706 (1)
1994	**0.0003 (0)**	**0.0009 (0)**	**0.0003 (0)**	**0.0045 (0)**	**0.0002 (0)**	**0.0033 (0)**
1995	0.9979 (1)	0.9994 (1)	0.9981 (1)	0.8503 (1)	0.9941 (1)	0.9993 (1)
1996	0.9978 (1)	0.9994 (1)	0.9980 (1)	0.8507 (1)	0.9941 (1)	0.9995 (1)
1997	0.9979 (1)	0.9994 (1)	0.9981 (1)	0.8531 (1)	0.9941 (1)	0.9995 (1)
1998	0.9957 (1)	0.9994 (1)	0.9965 (1)	0.8483 (1)	0.9939 (1)	0.9990 (1)
1999	**0.1303 (0)**	**0.1080 (0)**	**0.1261 (0)**	**0.1936 (0)**	**0.3335 (0)**	**0.2378 (0)**
2000	0.9979 (1)	0.9994 (1)	0.9980 (1)	0.8529 (1)	0.9941 (1)	0.9995 (1)
2001	**0.0003 (0)**	**0.0001 (0)**	**0.0001 (0)**	**0.0045 (0)**	**0.0003 (0)**	**0.0019 (0)**

The data set of 70 units is divided to 4 parts, the half of the data set is used for training, the quarter of the data set is used for validation and the remained quarter is used for testing. GDM, SCG and RP algorithms are applied as ANN algorithms. Consequently, although ANN models constructed by using the whole data give successful result, they give unsuccessful result for extrapolation. However, ANN models constructed by division of the data set can make successful extrapolation. Extrapolation for the data consist of 1988-2001 related to Turkey is given in Table 2.

Results Part 2

The 4:8:1 and 4:11:1 architectured ANN are selected from variety of experiments. The obtained correct classification ratios are summarized in Table 3. 4:8:1 and 4:11:1 with one hidden layer model which give better performance are selected as feed forward ANN architecture. The correct classification ratio for each algorithm is calculated and results are given in Table 4.

Comparisons are made according to correct classification ratios calculated by using probabilistic values related with the results of logistic and probit regression analysis of statistical techniques and ANN algorithms. It is obtained as a conclusion that ANN models give better results in classification. While SCG of ANN algorithms is giving more successful correct classification ratio, logistic regression analysis gives more successful results than probit regression.

ANN models constructed by using whole data give unsuccessful results for extrapolation. This situation can be explained as follows; ANN models constructed for whole data memorize the data, hence they don't give successful result for extrapolation. However, ANN models constructed by the divided data (training-validation-test) give successful result for extrapolation. If the results of the application are examined, it can be seen that only the target results for 1994, 1999 and 2001 can not be obtained. The reason can be explained by the economic crisis in these years.

Table 3. Results of ANN models (whole data) and Logistic, Probit Regression

Model / Criteria for Goodness of fit	SCG		GDM		RP		Lojistic Regression	Probit Regression
	4:8:1	4:11:1	4:8:1	4:11:1	4:8:1	4:11:1		
Correct Classification Ratio (%)	95.7	97	90	95.7	92.9	94.3	90	88.6

Table 4. Results of ANN models (training-validation-test)

Model / Criteria for Goodness of fit	SCG		GDM		RP	
	4:8:1	4:11:1	4:8:1	4:11:1	4:8:1	4:11:1
Correct Classification Ratio (%)	87	85.7	82.9	84.3	82.9	85.7

Clustering actual and candidate members of EU due to economical indicators

Countries are different from each other due to economical indicators such as inflation, budget deficit, gross domestic product etc. In order to construct monetary union, countries try to minimize their economical differences. For example, the Economic and Monetary Union (EMU) is one of the constructed unions. Studies on optimality (based on the similarity of economic indicators) of currency areas have started by the article published by Mundell in 1961 and followed by various studies. Since then, it is general investigation topic whether euro area countries or non-euro area countries satisfy the Maastricht Criteria or not.

Definition of the problem of clustering countries

Maastricht criteria describe the necessary conditions that actual member of EU have to satisfy for EMU entry, according to the Maastricht Treaty of February 7th, 1992.The Maastricht criteria are the five conditions set that countries had to meet if they want to take part in full economic and monetary union. These are:

1. Inflation - no more than 1.5% above the average inflation rate of the lowest 3 inflation countries in the EU

2. National debt - no more than 60% of GDP

3. Budget deficit - no more than 3% of GDP

4. Interest rates - the long-term rate should be no more than 2% above the average of the three countries with the lowest inflation rates

5. Exchange rates - currency within the normal bands of the ERM with no re-alignments for at least 2 years

Only 12 actual members of EU within 25 members are in the EMU and are in the euro area. These members are as follows: EU countries in euro area are Austria, Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Portugal, Spain and Holland. EU countries in non-euro area are Cyprus, Czech Republic, England, Estonia, Finland, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia and Sweden.

For evaluation of 25 actual members and 4 candidate members of EU based on Maastricht criteria for EMU, 4 economical indicators (inflation rate, budget deficit, interest rate, public debt) are taken into account.

Additionally 3 different variables (total employment rate, share of employment of corresponding country versus EU, share of external trade of corresponding country versus EU) are considered and the variables for data set are defined as follows:

X_1: Total Employment Rate (%)
X_2: Inflation Rate
X_3: Interest Rate

Table 5. Clusters by Ward's cluster technique

Cluster No	Countries in the cluster
1	Greece, Italy, Hungary, Slovenia, Malta, Portugal
2	Belgium, Sweden, Lithuania, Czech Republic, Germany, Estonia, France, Luxembourg Latvia, Ireland, Spain
3	Austria, Finland, Cyprus, Slovakia, Denmark, Poland, Holland, England

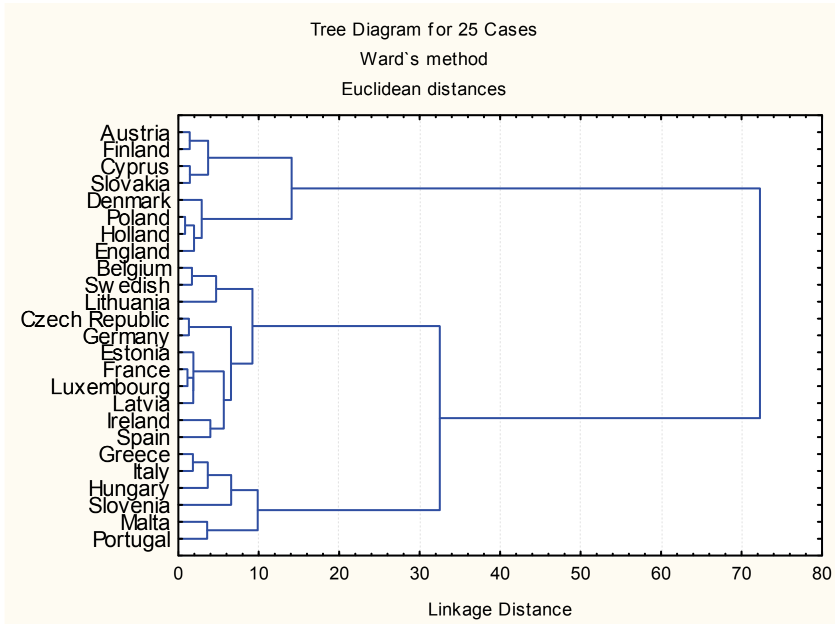

Figure 3 - Cluster dendrogram obtained by Ward's cluster technique

X_4: Budget Deficit/GDP
X_5: Public Debt/GDP
X_6: Share of employment
X_7: Share of external trade

(The data set of economical variables is collected from the EUROSTAT of 2003.)

Kohonen Nets of Neural Network Architectures (Kohonen, 1990) is recently a new cluster technique. This technique groups the data such as holding similar properties. Based on this property of Kohonen nets, in first stage this net is applied to cluster 25 actual members of EU and besides traditional statistical techniques of cluster analysis are applied. Then, in order to determine (classify) which set the 4 candidate members of EU belong, multilayer ANN techniques are applied. In this part of study, depending on the obtained results, which of those 29 countries can really join to EMU under the certain circumstances is identified and recommended. Applications are realized by using STATISTICA 6.0.

Analysis

In order to separate 25 member countries of EU in homogenous sets based on main macro economical indicators, use Ward's cluster technique of Hierarchical cluster analysis and then apply the k-means technique. Sets obtained by Ward's cluster technique are given in Table 5. A dendrogram obtained by Ward's cluster technique is given in Figure 3.

In this part of the application, K-means technique is applied to check the correctness of the clusters. The outcome of this technique is the same as the outcome of Ward's technique.

Kohonen Nets of ANN are applied to the same data in order to cluster the countries. To realize clustering using Kohonen nets the number of sets is given before as k-means technique. Because three sets are obtained by Ward's Clustering Technique, the same number of set is used to apply Kohonen Nets. After the training Kohonen Net, SOFM7:7-3:1 ANN model is obtained.

According to the win frequencies of this model, there are 6 countries in the first cluster, 7 countries in the second cluster and 12 countries in the third cluster. Observing the topological map of appropriate Kohonen Net, the clusters of the countries are given in Table 6.

After clustering 25 actual member of EU based on Ward's technique and K-means technique, at first, in order to determine (classify) which cluster the 4 candidate members of EU belong, multilayer ANN techniques are applied. MLP 6:6-6-3:1 model is selected as appropriate model within 1000 simulation. According to this model, clustering of the candidate countries of EU is given in Table 7.

Later, depending on cluster analysis by Kohonen Net, in order to determine (classify) which cluster the 4 candidate members of EU belong, multilayer ANN techniques are applied. RBF 6:6-3-3:1 is selected as appropriate model within 1000 simulation. According to this model, clustering of the candidate countries of EU is given in Table 8.

Results

Traditional statistical techniques of cluster analysis and Kohonen Nets of ANN are applied for clustering 25 actual members of EU. By each of the techniques, three separate clusters are performed, and especially it is seen that the third cluster of the both techniques consists of approximately the same countries. It is investigated that in which group of 4 candidate countries will take place according to basic macro economical indicators of 25 EU countries, and as another result of both of the techniques Croatia, Romania and Turkey take place in the same cluster and Bulgaria takes place in another cluster.

Table 6. Clustering countries using Kohonen Net

Cluster No	Countries in the Cluster
1	Greece, Hungary, Malta, Portugal, Slovenia, Spain
2	Cyprus, Czech Republic, Estonia, Italy, Latvia, Lithuania, Sweden
3	Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Luxembourg, Poland, Slovakia, Holland, England

Table 7. Clustering of the candidate countries of EU based on Ward's technique due to ANN

Cluster No	Countries in the cluster
2	Bulgaria
3	Croatia, Romania, Turkey

Table 8. Clustering of the candidate countries of EU based on Kohonen Net due to ANN

Cluster No	Countries in the cluster
1	Croatia, Romania, Turkey
3	Bulgaria

The unique euro-area country that satisfies all of the Maastricht Criteria in 2003 is Portugal. By 2003, actual members of EU that don't entry EMU but satisfy all of the Maastricht criteria are Czech Republic, Finland and Lithuania. Other countries' currency is Euro although they don't satisfy the criteria for that data set but they used to satisfy the criteria since 1993. On the other hand, with regard to some countries' not using Euro, although they satisfy the criteria, they can be explained by the political reasons.

According to traditional statistical techniques of cluster analysis, by 2003, countries in the first cluster satisfy the 3rd and 4th criteria, countries in second end third clusters satisfy at least one criterion. Bulgaria is a single candidate country of EU that mostly satisfies the Maastricht criteria, except for 1st criteria.

Conclusion

It is seen that for forecasting, classification and clustering problems, when suitable ANN algorithms are chosen better estimates are obtained compared to the statistical methods. Also, ANN has a great advantage of not requiring the assumptions which are necessary for statistical methods. Moreover, ANN has another advantage for extraneous data and prediction. However, the choice of suitable architecture and algorithm has a great importance. Depending on the idea of getting better results using RP and SCG algorithms for classification problems, those algorithms are suggested for that kind of problems. It is also necessary to be cautious about the QN algorithms that can give good results at the stage of choosing the suitable network algorithms. On the other hand, for the problem related with time series, it is seen that the hybrid models with both components are statistical or ANN models, give better performances. Also, the hybrid model's good performance with both nonlinear components changes the notion that the first component is linear.

Although in extrapolation problems when RFBN is applied by itself it is found unsuccessful, when it is used as second component of hybrid model, it provides better performance. In our opinion the reason of that is, after applying the first component to the data set, the difference between the units of neighborhood of the residuals lessens. That provides the opportunity of accumulation of residuals at a defined central (radial) points' neighborhood.

The result of being the hybrid model's more advantageous for forecasting problems; it gives the idea of constructing and using suitable hybrid models for other statistical problems.

The best model should be chosen depending on the statistical and the mathematical structure of the data set. Unfortunately, there aren't certain criteria or test about the selection of appropriate model subject to nature of the data. The numerous experimental studies which will be done on this subject can provide a definite and useful aggregation on a criterion.

References

Ansuj, A. P., Camargo, M. F., Radharamanan, R. & Petry, D. G. (1996). Sales forecasting using time series and neural networks. *Comput. Ind. Eng. 31* (1), 421-424.

Aslanargun, A., Mammadov, M., Yazici, B. and Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series tourist arrival forecasting. *Journal of Statistical Computation and Simulation*, Vol.77, Issue 1, 29-53.

Bhaya, A. & Kaszkurewicz, E. (2004). Steepest descent with momentum for quadratic functions is a version of the conjugate gradient method. *Neural Networks, 17*, 65-71.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Clarenden Press.

Box, G. P. & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco, CA: Holden-Day.

Caire, D., Hatabian, G., & Muller, C. (1992). Progress in forecasting by neural networks. *International Conference on Neural Networks II*, 540-545.

Chin, K. & Arthur, R. (1996). Neural network vs. conventional methods of forecasting. *Journal of Business Forecasting, 14* (4), 17-22.

Cho, V. (2003). A comparison of different approaches to tourism arrival forecasting. *Tourism Management, 24*, 323-330.

Cooper, J. C. B. (1985). Country creditworthiness: A canonical correlation analysis, *Glasgow Caledonian University Economics Discussion Paper*, No. 3.

Cooper, J. C. B. (1999). Artifical neural networks versus multivariate statistics: An application from economics, *Journal of Applied Statistics, 26*, 909-921.

Dhonte, P. (1975). Describing external debt situations: A roll over approach, *International Monetary Fund Staff Papers*, *24*, 159-186.

Feder, G. & Just, R. (1977). A Study of debt servicing capacity applying logit analysis. *Journal of Development Economics, 4*, 25-38.

Frank, C. R. & Cline, W. R. (1971). Measurement of debt service capacity: An application of discriminant analysis. *Journal of International Economics*, *1*, 327-344.

Ginzburg, I. & Horn, D. (1994). Combined neural networks for time series analysis. *Advancement in Neural Information Processing Systems, 6*, 224-231.

Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design*. Boston, MA: PWS Publishing.

Haykin, S. (1999). *Neural networks: A comprehensive foundation.* Prentice Hall Publishing Co.

Hill, T., O'Connor, M., & Remus, W. (1996). Neural network models for time series forecasts. *Management Science, 42*(7), 1082-1092.

Kenen, P. B. (1969). The theory of optimal currency areas: An eclectic view, in Mundell, R. A. & Swoboda, A. K. (eds*). Monetary Problems of the International Economy*, Chicago: University of Chicago Press.

Kharas, H. (1984). The long run creditworthiness of developing countries: Theory and practice. *Quarterly Journal of Economics*, *99*, 415-439.

Kim, T. Y., Oh, K. J., Sohn, I., & Hwang, C. (2004). Usefulnes of artifical neural networks for early warning system of economic crisis. *Expert Systems with Applications, 26*(4), 583-590.

Kim, J., Wei, S., & Ruys, H. (2003). Segmenting the market of West Australian senior tourists using an artificial neural network. *Tourism Management, 24*, 25-34.

Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series model for forecasting commodity prices. *Neurocomputing, 10* (2), 169-181.

Krugman, P. R. (1990). Policy problems of a monetary Union, in de Grauwe, P. & Papademos, L. (eds). *The European monetary system in the 1990s*. Harlow: Longman, (reprinted in P Krugman (ed) *Currencies and crises*. Cambridge, M.S.: MIT Press, 1992).

Kuan, C. M. & White, H. (1994), Artificial neural networks: An economic Perspective. *Economic Reviews, 13* (1), 1-91.

Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management, 21*, 331-340.

Law, R. & Au, N. (1999). A neural networks model to forecast Japanese demand for travel to Hong Kong. *Tourism Management, 20* (1), 89-97.

Leung, M. T., Chen, A., & Daouko, H. (2000). Forecasting exchange rates using general regression neural networks. *Computers and Operations Research, 27*(11-12), 1093-1110.

Luxhoj, J. T., Riis, J. O., & Stensballe, B. (1996). A hybrid econometric-neural-network modelling approach for sales forecasting. *Int. J. Prod. Econ. 43*, 175-192.

Maier, H. R. & Dandy, G. C. (1996). Neural network models for forecasting univariate time series. *Neural Networks World, 6* (5), 747-772.

McKinnon, R. I. (1963). Optimum currency areas. *American Economic Review*, *53*, 717-725.

Moller, M. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks, 6*(4), 525-533.

Mundell, R. A. (1961). A theory of optimum currency areas. *American Economic Review*, *51*, 657-64.

Nocedal, J. & Wright, S. J. (1999). *Numerical optimization*. New York, N.Y.: Springer-Verlag Publishing.

Pelikan, E., Groot, C. D., & Wurtz, D. (1992). Power consumption in West-Bohemia: Improved forecasts with decorrelating connectionist networks. *Neural Networks World,* 2, 701-712.

Qian, N. (1999). Efficient backpropagation learning using optimal learning rate and on the momentum term in gradient descent learning algorithms. *Neural Networks, 102*(1), 145-151.

Roy, J. & Cosset, J. (1990). Forecasting country risk ratings using a neural network. *Proceedings of the 23rd Hawaii International Conference on Systems Sciences*, *4*, 327-334 (Alamitos, CA, IEEE Pres).

Sharda, R. (1994). Neural networks for the MS / OR analyst: An application bibliography. *Interfaces, 24* (2), 116-130.

Su, C. T., Tong, L. I., & Leou, C. M. (1997). Combination of time series and neural network for reliability forecasting modeling. *J.Chin. Inst. Ind. Eng. 14* (4), 419-429.

Tseng, F. M., Yu, H. C., & Tzeng, G. H. (2002). Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting & Social Change, 69*, 71-87.

Voort, V. D., Dougherty, M., & Watson, M. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transp. Res. Circ. (energe Technol.), 4C* (5), 307-318.

Wang, J. H. & Leu, J. Y. (1996). Stock market trend prediction using ARIMA-based neural networks. *IEEE Int. Conf. Neural Networks, 4*, 2160-2165.

Wedding, D. K. & Cios, K. J. (1996). Time-series forecasting by combining RBF networks, certainty factors, and the Box-Jenkins model. *Neurocomputing, 10*, 149-168.

Wilson, R. & Sharda, R. (1992). Neural networks, *OR / MS today*. 36-42.

Wong, B., Bodnovich, T., & Selvi, Y. (1995). A bibliography of neural networks business application research. *Expert Systems, 12* (3), 253-262.

www.worldbank.org/data/countrydata

www.die.gov.tr

Yoon-Dae, E. (1979). *Commercial banks and the creditworthiness of less developed countries*. Ann Arbor, MI: UMI Research Press.

Yu X., H. & Chen, G. A. (1997). Efficient backpropagation learning using optimal learning rate and momentum. *Neural Networks, 10* (3), 517-527.

Zhang, G. P., Patuwo, E. B., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*, 35-62.

Zhang, G. P., Patuwo, E. B., & Hu, M. Y. (2001). A simulation study of artificial neural networks for nonlinear time-series forecasting, *Computer & Operations Research, 28*, 381-396.

Zhang, G. P. (2003). Time-series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159-175.

Zhang, G. P. & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*, 501-514.

Zhang, G. P. (2001). An investigation of neural networks for linear time-series forecasting. *Computer & Operations Research*, *28*, 1183-1202.