Jurnal Teknologi, 40(B) Jun. 2004: 1–16

UNIVARIATE ARTIFICIAL NEURAL NETWORK IN FORCASTING DEMAND OF LOW COST HOUSE IN PETALING JAYA

NORHISHAM BAKHARY[1], KHAIRULZAN YAHYA[2] & NG CHIN NAM[3]

Abstract. Recently researchers have found the potential applications of Artificial Neural Network (ANN) in various fields in civil engineering. Many attempts to apply ANN as a forecasting tool has been successful. This paper highlighted the application of Time Series Univariate Neural Network in forecasting the demand of low cost house in Petaling Jaya district, Selangor, using historical data ranging from February 1996 to April 2000. Several cases of training and testing were conducted to obtain the best neural network model. The lowest Root Mean Square Error (RMSE) obtained for validation step is 0.560 and Mean Absolute Percentage Error (MAPE) is 8.880 %. These results show that ANN is able to provide reliable result in term of forecasting the housing demand based on previous housing demand record.

Keywords: Time Series Univariate Neural Network, low cost housing demand, RMSE, MAPE

Abstrak. Kebelakangan ini ramai penyelidik mendapati '*Artificial Neural Network*' (ANN) untuk digunakan dalam berbagai bidang kejuruteraan awam. Banyak aplikasi ANN dalam proses peramalan menghasilkan kejayaan. Kajian ini memfokuskan kepada penggunaan siri masa 'Univariate Neural Network' untuk meramalkan permintaan rumah kos rendah di daerah Petaling Jaya, Selangor. Dalam kajian ini, beberapa kes bagi sesi latihan dan ramalan telah dibuat untuk mendapatkan model terbaik bagi meramalkan permintaan rumah. Nilai RMSE yang paling rendah yang diperolehi bagi tahap validasi adalah 0.560 dan nilai MAPE yang diperolehi adalah 8.880%. Hasil kajian ini menunjukkan kaedah ini memberikan keputusan yang boleh diterima dalam peramalan permintaan rumah berdasarkan data masa lalu.

Kata kunci: *Univariate Neural Network*, permintaan rumah kos rendah, RMSE, MAPE

1.0 INTRODUCTION

There has been a growing interest in application of neural computing in recent years for its capabilities, which include the ability to learn and generalize from the example, to produce meaningful solutions to problems even when the input data contains error or are incomplete, and to adapt solution overtime to compensate for the changing

[1] Lecturer, Faculty of Civil Engineering, Universiti Teknologi Malaysia, Skudai Johor. E-mail: Norhisham@utm.my
[2] Lecturer, Faculty of Civil Engineering, Universiti Teknologi Malaysia, Skudai Johor
[3] Former undergraduate student at Faculty of Civil Engineering, Universiti Teknologi Malaysia, Skudai Johor

situation [1]. The rapid advancement of neural computing and computer technology has also contributed to applications of artificial neural networks in science and technology. These new developments reinvigorated the field of artificial neural networks. In the last ten years, many papers have been written and artificial neural network (ANN) has found many applications in various engineering fields including civil and structural engineering. In construction, ANN has covered wide range of topics such as estimating costs and markup estimation [2], and predicting construction productivity [3]. [4] applied ANN as a tool for predicting the changes in construction cost indices.

According to Eight Malaysian Plan (2001-2005), about 232,000 units of low cost houses are required to be constructed during that period [5]. Due to the urbanization and increment on affordable houses every year, the government has carried out numerous campaigns and efforts in providing adequate shelters to accommodate the growing population, especially to cater for the low-income group. However, housing is one of the main aspects of urban problems which is directly linked to the economy. According to [6], shelter poverty which is caused by increasing housing costs relative to household income normally increases during an economic crisis. The experience of the 1987 property market recession had similar characteristics with the one experienced in 1997. The economic recession was followed by a fall in effective demand for properties; oversupply of residential, commercial and industrial buildings, and stock market crash in October 1987 [7]. This phenomenon exhibits that the housing demand fluctuates over time and it is difficult to accurately forecast the needs of houses.

There have been several attempts in predicting housing demands using computational methods. [8] utilized artificial neural network to forecast the residential construction demand in Singapore. It was concluded that ANN was able to give the best model compared to regression and multi-regression models. In the United States, Aiken, [9] reported that applying ANN model for forecasting semi-annual private residential houses is much better than the multi-linear regression model. [10] applied ARIMA model and ANN for predicting housing demand in Petaling Jaya, Selangor. They applied several relational factors as the input to forecast the demand of housing using multivariate neural network. It was concluded that the lack of information had an effect on the output pattern.

This study is an extension of a study by [10], where univariate neural network is utilised to forecast the housing demand. The objective of this study is to develop the best model for forecasting low cost housing demand in Petaling Jaya district.

2.0 TIME SERIES ANALYSIS

Time series analysis as described by most textbooks, relies on explicit descriptive, stochastic, spectral, or other methods of processes that describe the real world phenomena in generating the observed data [11].

Usually, the parameters of a standard model like ARIMA technique are derived from auto-correlation and frequency spectrum of the time series. One of the problems with ARIMA approach occurs when the time series represents non-linear processes. The usage of artificial neural network for time series analysis relies purely on the observed data. According to [12], multilayer feed forward networks with at least one hidden layer and a sufficient number of hidden neurons are capable of approximating any measurable function between input and output. The capability to generalize allows artificial neural networks to learn even in the case of noisy or missing data. Another advantage of a linear model like ARIMA technique is the network's ability to represent non-linear time series.

Basically, there are two types of time series forecasting models, namely the univariate and multivariate. Univariate models like Box Jenkins, use only one variable as the input to build the non-linear relationship between the input and output. The non-linear relationship will be used to forecast the desired future output.

For this study, fully connected, feed forward artificial neural network with one hidden layer and back-propagation learning algorithm is applied.

3.0 INTRODUCTION TO ARTIFICIAL NEURAL NETWORK

The foundations of backpropagation method for learning in neural network were explained by [13]. Artificial neural networks consist of many simple processing devices (called processing elements or neurons) grouped in layers. Each layer is identified by index $l = 0,\ldots,L$. The layers 0 and L are called the "input layer" and "output layer" respectively. Other layers are called the "hidden layers". Communication between processing elements is only allowed for processing elements of neighbouring layers. Neurons within layers cannot communicate. Each neuron has a certain activation level, a. The network processes data by the exchange of activation level between connected neurons (see Figure 1)

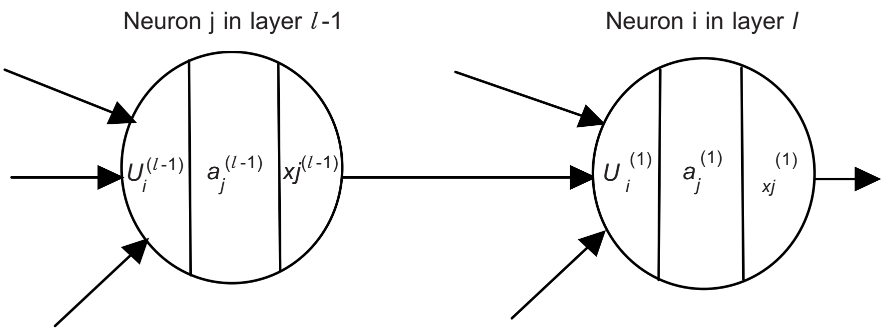

Figure 1 Exchange of activation values between neurons

The output value of the i-th neuron in layer l is donated by $x_i^{(l)}$. It is calculated with the formula:

$$x_i^{(l)} = g\left(a_i^{(l)}\right) \tag{1}$$

where $g(\ )$ is the monotone increasing function. In this case, sigmoid function $\left(g(y) = \dfrac{1}{1+e^{-y}}\right)$. The activation level $a_i^{(l)}$ of the neuron i in the layer l is calculated by

$$a_i^{(l)} = f\left(u_i^{(l)}\right) \tag{2}$$

where $f(\ )$ is the activation function.

The net input $u_i^{(l)}$ of neuron I in the layer l is calculated as:

$$u_i^{(l)} = \left(\sum_{j=1}^{n(l)} w_{ij}^{(l)} x_j^{(l-1)}\right) - \oplus_i^{(l)} \tag{3}$$

where $w_{ij}^{(l)}$ is the weight of neuron j in layer l–1 connected to neuron i in layer l, and $x_j^{(l-1)}$ is the output of neuron j in layer l–1. $\oplus_i^{(l)}$ is a bias value that is subtracted from the sum of the weighted activations.

The calculation of the network status starts at the input layer and end at the output layer. The input vector I initializes the activation levels of the neurons in the input layer:

$$a_i^{(0)} = i^{th} \text{ element of } I \tag{4}$$

For the input layer, g is the identity function. The activation level of one layer is propagated to the next layer of the network and the weights of the neurons are changed by the backpropagation learning rule.

One of the most popular and widely implemented of all neural network paradigms is backpropagation algorithm. It is a systematic approach for training multiple layer of neural network [14]. Using backpropagation algorithm, two propagation passes forward and backward are required.

There are two main steps in developing neural network model where training and testing process are required (sometimes validation steps applied). Training process starts by presenting a set of data to the network called the training set. The training sets consist of input and desired output. The network will try to learn the pattern from the given data. Then, testing steps are applied after the training process. A new set of data which has not been introduced to the present neural network model is applied and provides neural network performance.

4.0 IMPLEMENTATION

Univariate time series forecasting analyses past data to estimate future values. Basically, this method is to model a nonlinear function by recurrence from past values. The recurrence relation can then be used to predict the new values in time series, which hopefully would be the good approximation of the actual values. Univariate models contain only one variable in the recurrence equation. In this study, past housing demand data was used as the variable to predict the values in the future.

The data used in this study consists of monthly housing demand record in Petaling Jaya, Selangor from February 1996 to November 2000 [15]. The data is divided into three parts i.e., for training, testing, and validation. A sufficient training would yield a better neural network prediction [14]. However, excessive training would result in "*overtraining*" where the neural network model tends to memorize the training data. This may result in a good training performance but poor prediction. Thus, the number of data used in this study for training was varied to five cases of different data distribution to identify the best trained network. Table 1 shows the distribution percentage for each case.

Table 1 Five cases of different data distribution for training, testing, and validation

Case	**1**	**2**	**3**	**4**	**5**
Percentage for training (%)	66.66	80	50	70	60
Percentage for testing and validation (%)	33.34	20	50	30	40

In order to ensure that the data will contribute evenly to the model, data normalization was applied. The natural log (Ln) function was used as the data normalization method before applying the data to neural network model. By using the Ln function, the data was transformed to smaller scale between 0 to 10.

The forecasting accuracy was objectively measured by comparing the mean absolute percentage error (MAPE) of all the cases between the forecasted demand and the actual demand from June 2000 to November 2000.

The neural network was modeled using Neural Connection software. Back-propagation algorithm was used as learning algorithm and sigmoid function was used as the activation function for the hidden layers. Figure 2 shows the architecture of the model used.

It is important to determine a suitable network architecture in ANN modelling. The architecture consists of the number of layers, the number of processing elements (nodes) in each layer, and the interconnection scheme between the layers in ANN. The optimum architecture is often achieved by trial and error according to the complexity of the respective problem and also by testing few proposed designs to select the one that gives the best performance. Although there are no specific rules

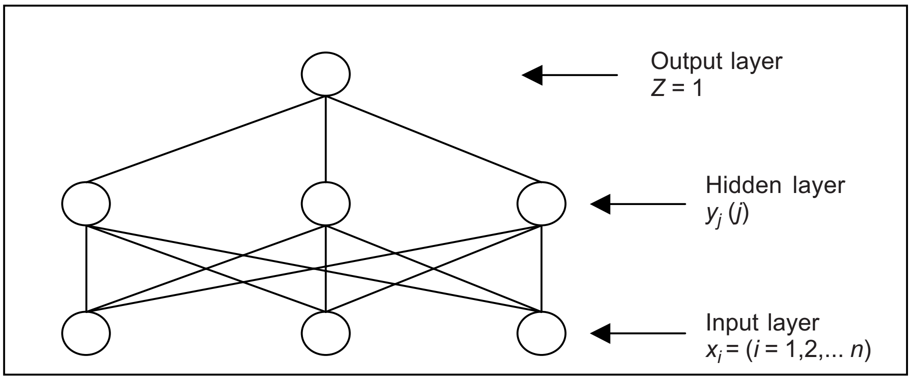

Figure 2 Neural network architecture used in the study

governing the design of ANN architecture, basic understanding of the functions on the different layers in the network helps to justify the choice of specifications. The input layer presents data to the network. The number of input nodes in time series univariate analysis is determined by the interval time based on the type of data used. For example, 12 input nodes are used for monthly series, 3 nodes for quarterly series, and so on.

One hidden layer has been used in this study. The output layer consists of one node that corresponds to the output variable demand data. In the training, steepest descent algorithm, and sigmoid activation function were applied. The number of hidden neurons was varied in every case, as shown in Table 2, and all software default setting parameters were used for training until the best combination was achieved.

5.0 RESULTS AND DISCUSSION

After several series of trial and error using five different cases of data distribution, the result of training for Case 5 with eleven hidden nodes depicting the lowest Root Means Square Error (RMSE) and the best Mean Absolute Percentage Error (MAPE), was chosen for forecasting.

Based on the results (Table 2), Case 5 was chosen for model forecasting because of the lowest RMSE (0.189) and the best value of MAPE (2.276%) in the training set. Figure 3 shows a close pattern between the actual and the trained data due to the ability of the network to train the 27 sets of data used in the training set. However, for testing, the output is unable to produce the lowest RMSE and MAPE. This may be due to overtraining problem in training set as very few the data was trained in the 800 range of housing demand. Furthermore, the 13 sets of data used in testing set failed to represent the overall population in network. Figure 4 shows a fairly similar pattern between the actual and the tested data in the testing set.

After the training and testing processes, Case 5 was taken as the best trained model. A set of data was introduced to the model for validation purposes. In this

Table 2 Neural network performance using different number of hidden neurons

Hidden nodes	Case	Training		Testing	
		RMSE	MAPE (%)	RMSE	MAPE (%)
11	1	0.227	2.495	0.689	9.819
18	2	0.560	4.000	0.342	5.302
11	3	0.281	3.512	0.597	7.566
19	4	0.313	2.928	0.689	9.767
11	5	0.189	2.276	0.647	8.927

Note: RMSE = Root Mean Square Error, MAPE = Mean Absolute Percentage Error

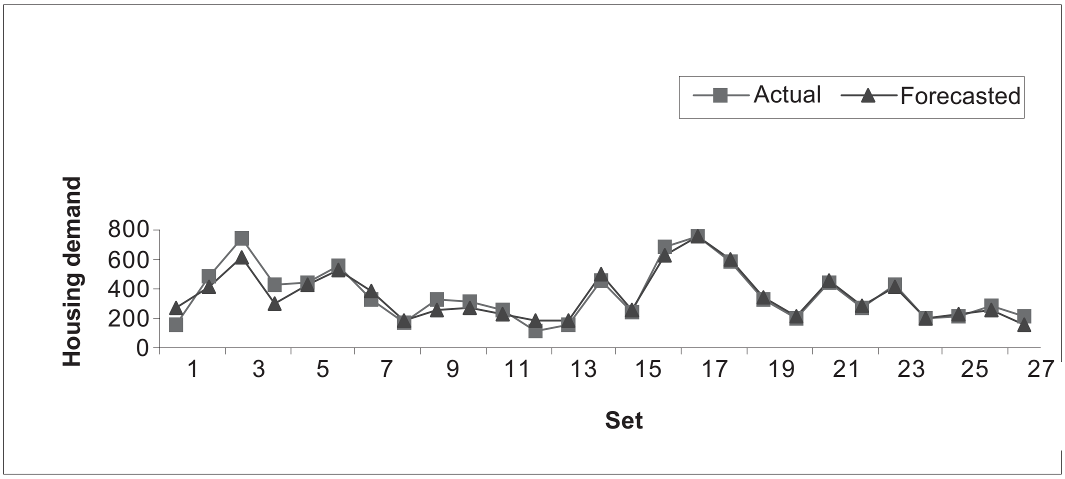

Figure 3 Forecasted and actual housing demand in training set

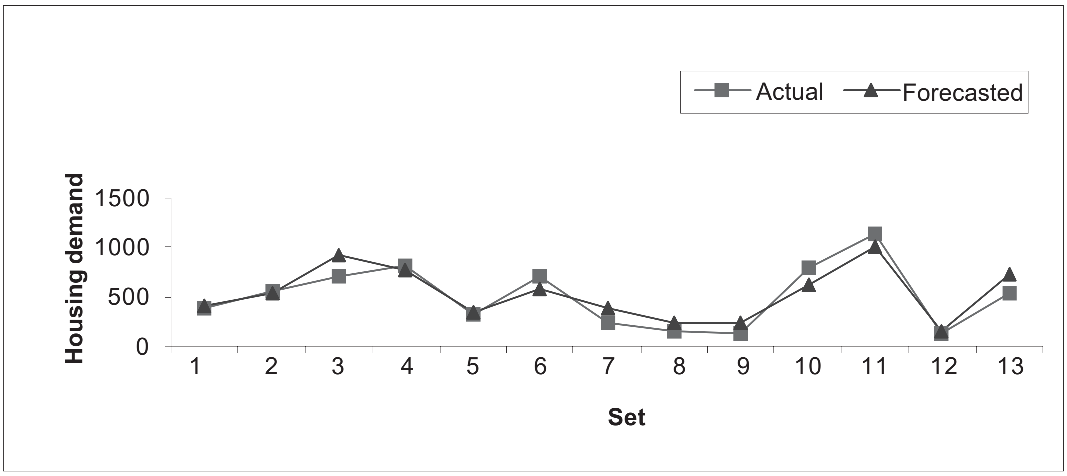

Figure 4 Forecasted and actual housing demand in testing set

study, the housing demand data from June 2000 to Nov 2000 was used to validate the model.

Figure 5 shows the plotted results of Case 5 compared to the actual housing demand from June 2000 to November 2000.

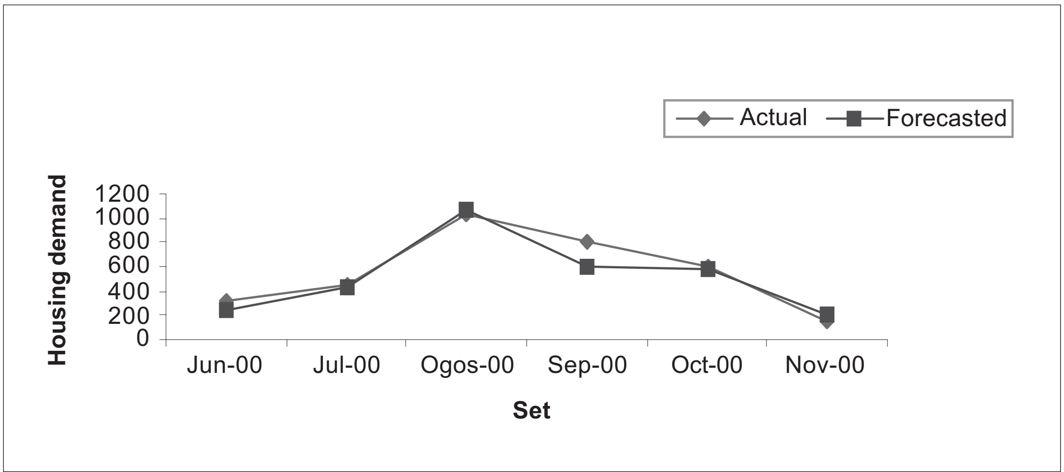

Figure 5 Forecasted and actual housing demand over 6 months ahead

Table 3 shows the summary of complete result for training, testing, and validation of the five cases.

Table 3 Results of training, testing, and validation sets for all cases

Case	Training		Testing		Validation	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
1	0.227	2.495	0.689	9.819	0.535	7.839
2	0.560	4.000	0.342	5.302	1.464	20.579
3	0.281	3.512	0.597	7.566	0.649	9.524
4	0.313	2.928	0.689	9.767	0.560	6.913
5	0.189	2.276	0.647	8.927	0.560	8.880

From the table above, it is observed that the forecasting pattern is relative to the data distribution used in network. Smaller percentage distribution of data (20%) for testing and validation sets produced higher RMSE and MAPE for the training set, as seen in Case 2. The optimum distribution was achieved when results of RMSE and MAPE in the training set is the lowest amongst all cases (Case 5). When the percentage distribution of data for testing and validation sets was higher than the optimum distribution, the RMSE and MAPE values in the training set tend to increase (see

Case 3). In this study, Case 5 was chosen for the final modelling because of the optimum distribution of data in training, testing, and validation sets.

6.0 CONCLUSION

Results of this study indicate that time series Univariate Artificial Neural Network (ANN) is able to forecast the housing demand with reliable accuracy based of Mean Absolute Percentage Error (MAPE) of less than ten percent. The results also show that the suitability of distribution and sizes of data used can affect the accuracy of the forecasting. The model can be further improved by increasing the size of data used in the ANN modelling and using other types of neural network such as radial basis network and recurrent network. It has been shown that univariate neural network has a potential to be used as a meaningful tool for the purpose of forecasting housing demand. The results also illustrates that the future values of housing demand are influenced by past and current housing demand data.

REFERENCES

[1] Flood, I., and N. Kartam. 1994. Neural Networks in Civil Engineering. I: Principal and Understanding, *Journal of Computing in Civi. Engineering. ASCE.* 8(2): 131-148.

[2] Moselhi, O., T. Hegazy., and P. Fazio. 1991. Neural Networks as Tools in Construction. *Journal of Construction. Engineering and Management. ASCE.* 117(4): 606-625.

[3] Chao, L. C., and M. J. Skibniewski. 1994. Estimating Construction Productivity: Neural-Network Approach. *Journal of Computing in Civil. Engineering. ASCE.* 8(2): 234-251.

[4] William, T. P. 1994. Prediction Changes In Construction Cost Indexes Using Neural Network. *Journal of Construction Engineering and Management*. 120: 306-319.

[5] Website: http://www.kpkt.gov.my/jabatan/jpn/artikel4.htm#table7. National Housing Department (NHD) 2001. Towards successful housing development in Malaysia.

[6] Stone, M. E. 1980. Housing and the American economy. In C. Pierre, .J. Forrester and W.W.Goldsmith (eds,). *Urban and Regional Planning in Age of Austerity*. New York: Pergamon.

[7] Malaysia. 1988. *Laporan Pasaran Hartanah 1987*. Kementerian Kewangan Malaysia: Percetakan Negara, Kuala Lumpur.

[8] Goh, B.H., 1998. Forecasting Residential Construction Demand in Singapore: A Comparative Study of The Accuracy of Time Series, Regression and Artificial Neural Network Techniques. *Journal of Engineering. Construction and Architectural Management*. 5: 261-275.

[9] Chartfield. E. 1991. *The Analysis of Time Series: Forth Edition*. New York: Chapman and Hall.

[10] Mohd Zaimi A.M., and Y. Khairulzan. 2002. Comparative Study on Forecasting Demand of Low Cost House in Urban Areas Using Artificial Neural Networks and Arima Model. *Conference Proceeding of 1st International Conference on Challenges and Opportunities in Management and Technology*. Miami. Florida. 2002.

[11] Aiken, M., B. Wailer., and T. Greer. 2001. Forecasting Construction Demand with a Neural Network. Technical Article. 1-4.

[12] Hornik, K., M. Stinchcombe., and H. White. 1989. Multilayer Forward Networks are Universal Approximators. *Neural Network*. 253-239.

[13] Rumelhart, D. E., and J. L. McClelland. 1986. *Parallel Distribution Processing*. Massachusetts: MIT Press.

[14] Lippmann, R. P. 1987. An Introduction to Computing with Neural Net. *IEE ASSP Magazine*. 4-22.

[15] Khairulzan, Y. 2002. *Penilaian Terhadap Keupayaan Model Rangkaian Neural Bagi Meramalkan Permintaan Rumah Kos Rendah Kawasan Bandar di Selangor*. Master Thesis. Universiti Teknologi Malaysia.