Estate Market Forecast Using Artificial Neural Networks

Dr. Qeethara K. Al-Shayea
Faculty of Economics and Administrative Sciences
Al Zaytoonah University of Jordan
Amman Jordan
kit_alshayeh@yahoo.com

International Conference on Business Intelligence and Knowledge Economy

Al Zaytoonah University of Jordan, Faculty of Economics and Administrative Sciences
23-26 (April 2012)
Amman Jordan

Abstract:

Artificial neural networks are widely used in business disciplines. The objective of this study is to provide independent real estate market forecasts on home prices using artificial neural networks. The Cascade Forward Back Propagation (CFBP) neural network is used to forecast house price, based on selected 13 parameters which are considered as forecast variables. The results of applying the CFBP neural networks methodology to forecast house price based upon selected parameters show abilities of the network to learn the patterns. In all cases, the percent correctly forecast in the simulation sample is above 94 percent. Empirical results support the potential of artificial neural network on house price forecast. CFBP neural networks are successfully used model for prediction, classification and forecasting.

Keywords: Cascade Forward Back Propagation (CFBP), Artificial Neural Network, Business Intelligence.

1. Introduction

An accurate prediction on the house price is important to prospective homeowners, developers, investors, appraisers, tax assessors and other real estate market participants, such as, mortgage lenders and insurers [2]. Traditional house price prediction is based on cost and sale price comparison lacking of an accepted standard and a certification process. Therefore, the availability of a house price prediction model helps fill up an important information gap and improve the efficiency of the real estate market [1].

Sean Zdenek [12] produced AI systems by rhetorical means; it does not merely describe AI systems or reflect a set of prevailing attitudes about technology

Kontrimas, Vilius; Verikas, Antanas [9] used the ordinary least squares (OLS) linear regression is the classical method used to build models in this approach. The method is compared with computational intelligence approaches – support vector machine (SVM) regression, multilayer perceptron (MLP), and a committee of predictors. The performance of the committee using the weights based on zones obtained from the SOM was also higher than of that exploiting the real estate value zones provided by the Register center.

Landers, Jay [10] produced offers information on the market forecasts for the nonresidential construction sector in the U.S. in 2008. A report titled "Construction Outlook 2008," by the company McGraw-Hill Construction says that more rigid lending standards have begun to affect commercial real estate and have reduced the volume of property purchases. According to a forecast by FMI Corp., there will be a marginally better result for total construction in 2008.

Kaihla, Paul, Copeland, Michael V., Hawn, Carleen, Lappin, Todd, Lev-Ram, Michal, Sloan, Paul [8] presented information related to the current status of real estate in the U.S. The nationwide housing slump was most evident with the fact that the median sales price for existing U.S. homes slipped to $225,000 in August 2006. It is however opined that the impact of housing downturn on the American residential real estate assets will not be much extensive.

Grudnitski, Quang Do and Shilling [4] applied a neural network analysis to supply evidence that answers this question. We find evidence that the characteristics of a borrower's net worth, marital status and education level and whether a co borrower is involved contribute in a significant way to the neural network's ability to determine mortgage choice.

2. Artificial Neural Networks

Neural network is an artificial intelligence model originally designed to replicate the human brain's learning process. The model consists of three main layers: input data layer (example the property attributes), hidden layer(s) (commonly referred as "black box"), and output layer (estimated value). Neural network is an interconnected network of artificial neurons with a rule to adjust the strength or weight of the connections between the units in response to externally supplied data (Fig. 1) [6].

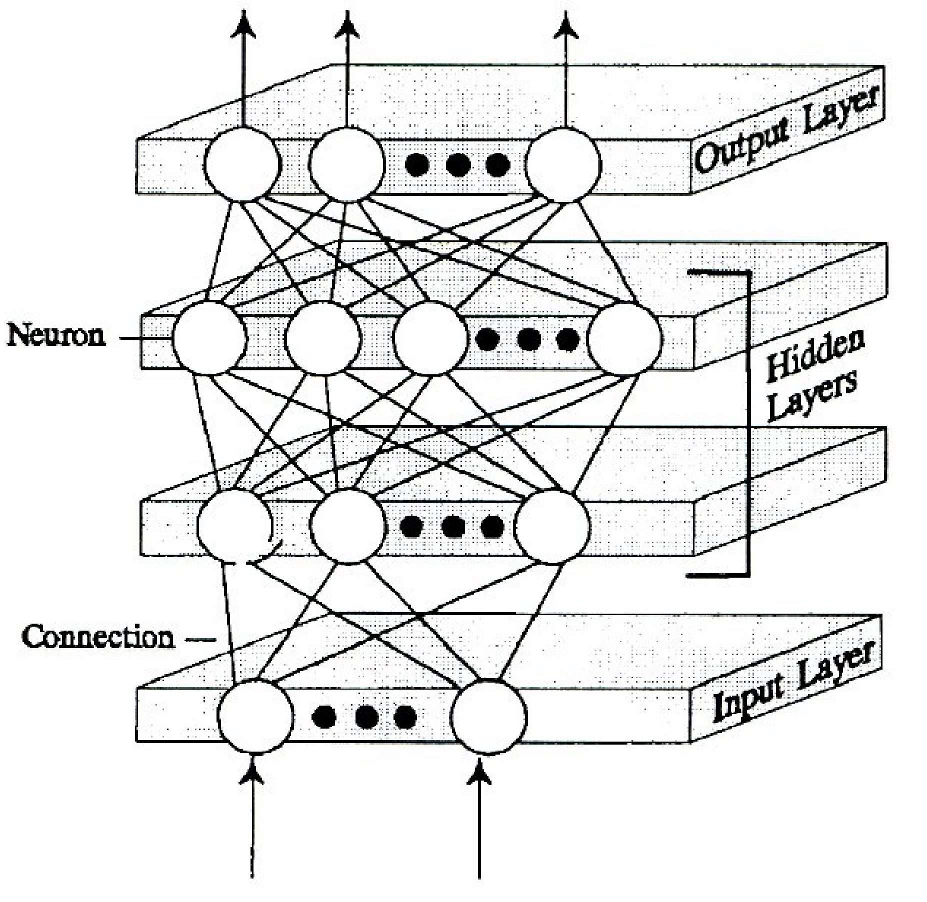

Fig. 1 A typical neural network

Each artificial neuron (or computational unit) has a set of input connections that receive signals from other computational units and a bias adjustment, a set of weights for input connection and bias adjustment, and transfer function that transforms the sum of the weighted inputs and bias to decide the value of the output from computational unit. The output for the computation unit (node j) is the result of applying a transfer function j to the summation of all signals from each connection (Ai) times the value of the connection weight between node j and connection i (Wji) (refer to equations 1 and 2) [6].
Sumj = Sj (Sum WjiAi)
Oj = j (Sumj)
where Oj is output for node j and j is transfer function which can take many different forms: linear functions, linear threshold functions, step linear functions, sigmoid function or Gaussian functions [6].

2.1 Cascade-Forward Back Propagation

The Cascade-forward networks. These are similar to feed-forward networks, but include a weight connection from the input to each layer and from each layer to the successive layers. For example, a three-layer network has connections from layer 1 to layers 2, layer 2 to layer 3, and layer 1 to layer 3. The three-layer network also has connections from the input to all three layers. The additional connections might improve the speed at which the network learns the desired relationship.

Cascade Forward Back Propagation (CFBP) is similar to FFBP network in using the BP algorithm for weights updating, but the main symptom of this network is that each layer neurons relates to all previous layer neurons. This kind of modeling problem usually requires a back propagation. We will use the back propagation network because it is by far the most popular neural network model and it is, in some respects, the easiest to work with. The architecture of the neural network is mostly subject to our data representation decisions. The number of inputs and the number of outputs are determined by these choices. Our major architectural decision deals with the number of hidden layers and hidden units.

The use of the neural network model is similar to the process utilized in building CFBP house price model. However, the neural network must first be trained from a set of data. For a particular input, an output (forecasted house price) is produced from the model. Then, the model compares the model output to the actual output (target output). The accuracy of this value is determined by the total mean square error and then back propagation is used in an attempt to reduce forecast errors, which is done through the adjusting of the connection weights. The performance of the network can be influenced by the number of hidden layers and the number of nodes that are included in each hidden layer. Therefore, a trial-and-error process is applied to find the optimal artificial neural network model. A Cascade-forward back propagation neural network software package was used to construct the artificial neural network model. There are no assumptions about functional form, or about the distributions of the variables and errors of the model, neural network model is more flexible than the standard statistical technique. It allows for nonlinear relationship and complex classificatory equations. The user does not need to specify as much detail about the functional form before estimating them classification equations but, instead, it lets the data determine the appropriate functional form.

3. Forecasting with Artificial Neural Networks

An artificial neural network consists of neurons, which have been related with special arrangement. Neurons are in layers and every network includes some neurons in input layer, one or more neurons in output layer and neurons in one or more hidden layers. Algorithms and architectures of artificial neural networks are different through variation in neuron model and relationship between neurons, and their weights. The learning purpose in artificial neural networks is weights updating, so that with presenting set of inputs, desired outputs are obtained. The most common types of artificial neural networks include: feed forward, feed back and competitive [11]. Training is a process that finally results in learning. Each network is trained with presented patterns. During this process, the connection weights between layers are changed until the differences between forecasted values and the target (experimental) is reduced to the permissible limit. Weights interpret the memory and knowledge of network. the advantages of using ANN are: high computation rate, learning ability through pattern presentation, prediction of unknown pattern and flexibility affront the noisy patterns [5].
ANN is trained such that a particular input leads to a specific target output. There are generally four steps in the training process: (1) assemble the training data, (2) create the network object, (3) train the network, and (4) compute the network response to new inputs. The Cascade Forward Back Propagation has 2 layers. Neural Network is trained for 13 epochs. The structure of the Cascade Forward Back Propagation network is shown in Fig. 2.

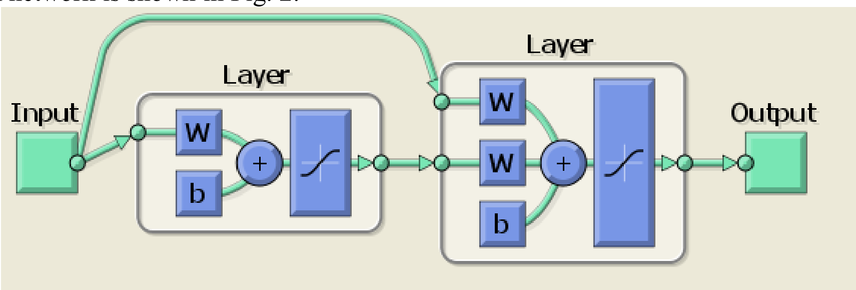

Fig. 2 The Structure of the Cascade Forward Back Propagation Network.

The first layer has 13 neurons;, and the output layer has 1 neuron. The size of the input vector is 13 x 506 and size of the target vector is 1x 506 in this structure. TANSIG is calculated according to [2]

$$Y_j = \frac{2}{(1+\exp(-2X_j))-1} \qquad \text{(TANSIG)}$$

And Mean square error is calculated according to [2]

$$MSE = \sum_{p=1}^{M} \sum_{i=1}^{N} (S_{ip} - T_{ip})^2$$

Where *R*2 is the determination coefficient, *mr E* the mean relative error, *mr SD* the standard deviation of mean absolute error, *k S* the network output for *k*th pattern, *k T* the target output for *k*th pattern and n the number of training patterns [2].

$$R^2 = 1 - \frac{\sum_{k=1}^{n}[S_k - T_k]}{\sum_{k=1}^{n}[S_k - \frac{\sum_{k=1}^{n} S_k}{n}]}$$

$$E_{mr} = \frac{100}{n}\sum_{k=1}^{n}\left|\frac{S_k - T_k}{T_k}\right|$$

$$SD_{mr} = \sqrt{\frac{\sum_{k=1}^{n}(\left|\frac{S_k - T_k}{T_k}\right| - \left|\frac{S_k - T_k}{T_k}\right|)}{n-1}}$$

In which $X j$ is the sum of weighed inputs for each neuron in *j*th layer and computed as below [2]:

$$X_j = \sum_{i=1}^{m} W_{ij} \cdot Y_i + b_j$$

Where m is the number of output layer neurons, $W ij$ the weight of between *i*th and *j*th layers, $Y I$ the *i*th neuron output and $b j$: bias of *j*th neuron for CFBP networks. As can be observed from the results, models with Cascade Feed Forward Back Propagation neural network structure gives the best results because mean square error value is less than those of others.

3.1 Data Analysis

The data was created by a house price as a data set to test the expert system, which will perform the forecasting system. The data is this dataset was taken from the StatLib library which is maintained at Carnegie Mellon University. The main idea of this data set is to construct the neural network model, which will perform the forecast house price. For better understanding of the problem let us consider definitions of house price. A house price is characterized by Median value of owner-occupied homes in $1000's as a target values. This dataset contains 506 Number of Instances. Table 1 presents the parameter of data which are considered as forecasting variables. The dataset contains 506 samples.

parameter of data (forecasting variables)	
No	**Parameters**
1	**per capita crime rate by town**
2	**proportion of residential land zoned for lots over 25,000 sq.ft.**
3	**proportion of non-retail business acres per town**
4	**Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)**
5	**nitric oxides concentration (parts per 10 million)**
6	**average number of rooms per dwelling**
7	**proportion of owner-occupied units built prior to 1940**
8	**weighted distances to five Boston employment centers**
9	**index of accessibility to radial highways**
10	**full-value property-tax rate per $10,000**
11	**pupil-teacher ratio by town**
12	**1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town**
13	**% lower status of the population**

3.2 Performance Evaluation

A Cascade-forward network with 13 inputs and 2 hidden neurons and linear output neurons was created using the neural network toolbox from Matlab 7.9. Inflammation of forecasted house price. Such net can fit multi-dimensional mapping problems arbitrarily well, given consistent data and enough neurons in its hidden layer as shown in Fig. 3. Training network automatically stops when generalization stops improving, as indicated by an increase in the mean square error (MSE) of the validation samples. The results of applying the artificial neural networks methodology to forecast between input data (actual data) and target data

The network was simulated in the testing set (i.e. cases the network has not seen before). The results were very good. Best validation performance is 14.294 at epoch 7 as shown in Fig. 3.

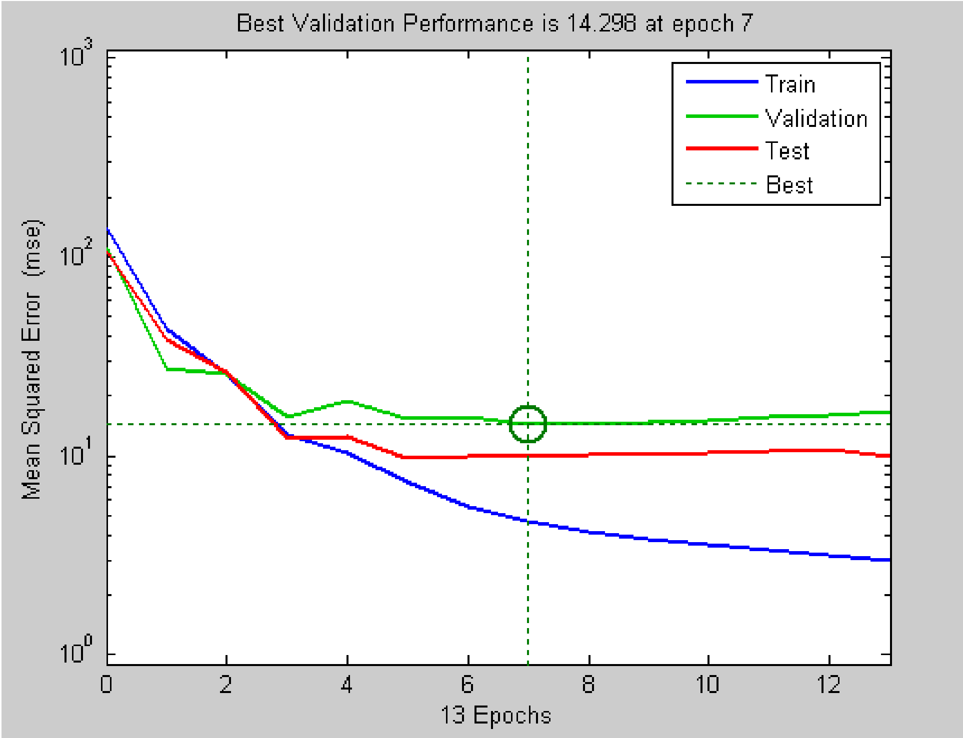

Fig. 3

Table 2 presents the values for training, validation and testing samples, and all three type of regression in the Fig 4.

Table 2: The Regression values for the training, validation and testing.

type	Regression value
Training	0.978
Validation	0.927
Testing	0.935
All type	0.96

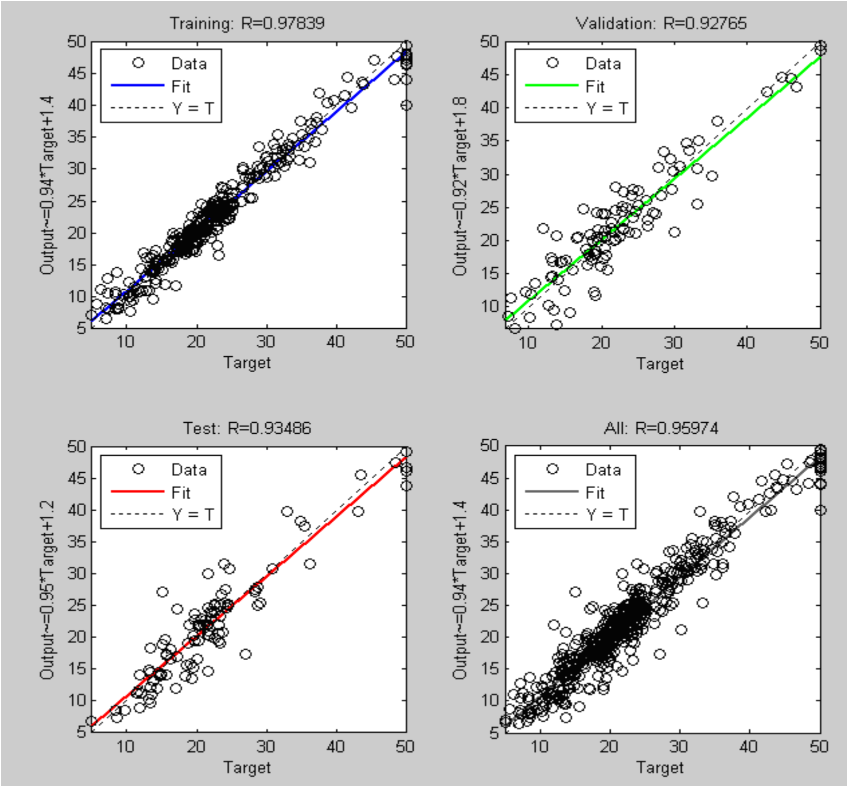

Fig. 4

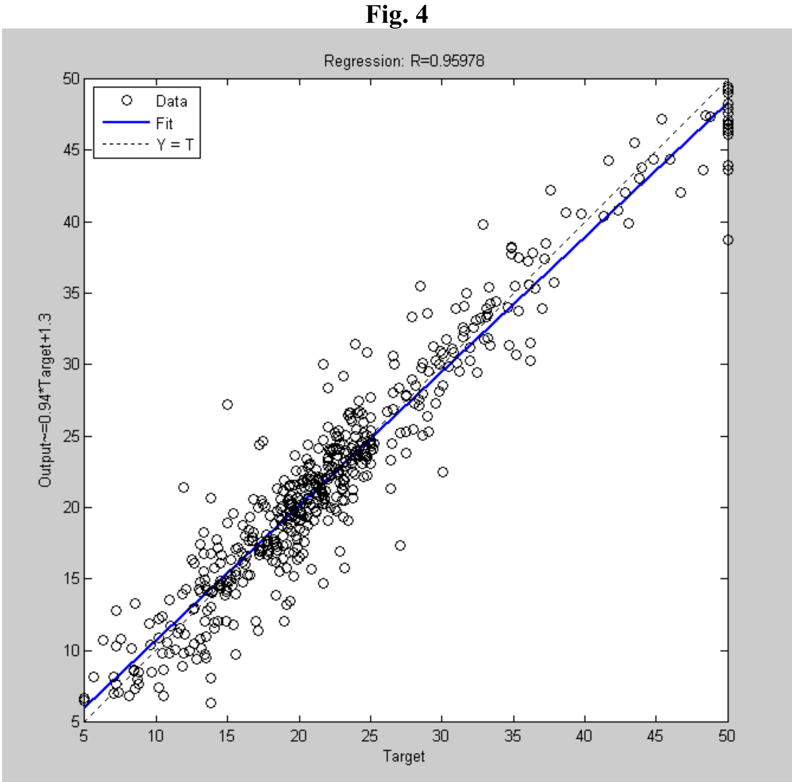

Fig. 5

The regression graph is get it from the scatter plot in Fig. 5 the relation between Output_ network is result from the equation output = 0.94 * Target +1.3, and the target value (Median value of owner-occupied homes in $1000's). The regression value is approximately 1 R = 0.96. That mean the output_ network is matching to the target data set (Median value of owner-occupied homes in $1000's), and the percent correctly forecast in the simulation sample is 96 percent.

4. Conclusion:

An accurate prediction on the house price is important to prospective homeowners, developers, investors, appraisers, tax assessors and other real estate market participants. The objective of this study is to provide independent real estate market forecasts on home prices using artificial neural network. The CFBP neural network is used to forecast house price. The results of applying the CFBP neural networks methodology to forecast house price based upon selected parameters show abilities of the network to learn the patterns. Artificial neural networks showed significant results on house price forecast in the simulation sample is above 96 percent.

5. Reference:

1- Calhoun, C. A., 2003. Property Valuation Models and House Price Indexes for the Provinces of Thailand: 2000. Housing Finance International, 17, pp. 31 – 41.

2- Demuth, H. and M. Beale. 2003. Neural Network Toolbox for Matlab- Users Guide Version 4.1. Natrick, U.S.A.: The Mathworks Inc.

3- Frew, J., and G. D. Jud, 2003. Estimating the Value of Apartment Buildings, The J. Real Estate Res., 25, pp. 77 - 86.

4- G. Grudnitski, A. Quang Do and J. D. Shilling, A Neural Network Analysis of Mortgage Choice, Intelligent Systems In Accounting. Finance and Management Vol. 4, pp. 127-135 (1995)

5- Heristev, R.M. 1998. The ANN Book GNU Public License, Available on
htttp://ftp.funet.fi/pub/sci/neural/books/

6- 35James, R. C. and E. B. Carol, 2000. Artificial Neural Networks in Accounting and Finance: Modeling Issues, Int. J. Intelligent Systems in Accounting. Finance & Management, 9: 119 – 44.

7- Jayanta K. Bandyopadhyay, and Bratati Bandyopadhyay, An Algorithm For Setting The List Price Of A Residential Real Estate Property, Department of Management, Central Michigan University, MaGraw Hall 105, Mount -1981.

8- Kaihla, Paul, Copeland, Michael V., Hawn, Carleen, Lappin, Todd, Lev-Ram, Michal, Sloan, Pau, The New Rules Of Real Estate, Nov 2006, Vol. 7, Issue 10.

9- Kontrimas, Vilius; Verikas, Antanas. The mass appraisal of the **real estate** by computational **intelligence**, Applied Soft Computing, Jan 2011, Vol. 11 Issue 1, pp443-448.

10- Landers, Jay, Market Forecasts See Mixed Conditions for Nonresidential Construction,Civil Engineering (08857024), Apr 2008, Vol. 78 Issue 4, pp14-16.

11- Menhaj, M.B., Fundamental of Artificial Neural Networks. Tehran, Iran: Amir Kabir University Publishing, 1998.

12- Sean Zdenek, Artificial intelligence as a discursive practice: the case of embodied software agent systems, October 2003 _ Springer-Verlag London Limited 2003.