House Rent Estimation in Dhaka City by Multi Layer Perceptions Neural Network

Samsuddin Ahmed ,Md.Mahbubur Rahman and Sabirah Islam

Dept. of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh
sambd86@gmail.com, mahabub.cse.buet@gmail.com, sabirahislam01@gmail.com

Abstract

This paper attempts to build an artificial neural network that can estimate the median price of a home in a neighborhood described by forty demographic attributes in areas of Dhaka City namely Dhanmondi, Baridhara, Gulshan, Mirpur, Uttara and Old Dhaka. A structured questionnaire was used to collect the relevant data and the housing data sets was used to develop constant quality price indices using traditional econometric techniques and using neural networks incorporating genetic algorithms. Factors including house size, house age, house type, number of bedrooms, number of bathrooms, number of garages, amenities around the house and geographical location are considered. The analysis indicates that neural networks act as real alternative to the econometric methods. In this paper thirteen conditional attributes have been considered to estimate the house rent. This paper provides some indicative policy guidelines to handle the house rent problem in the Dhaka City and suggests that a rent controller should be appointed for each ward and maximum rents for particular areas should be gazette and the rents should be paid through banks.

Keywords: Accessibility, House rent estimation, Neural Network

1. Introduction

An accurate prediction on the house rent is important to prospective homeowners, developers, investors, appraisers, tax assessors and other real estate market participants, such as, mortgage lenders and insurer. Therefore, the availability of a house rent prediction model helps fill up an important information gap and improve the condition of general people. Dhaka is the nucleus city of Bangladesh and has come to be known as one of the densely megacity in the world. The city in 2013 had over 20 million people for the larger conurbation and 8 million people within the central city area. The quick growth of population of Dhaka has been caused by high rate of in-migration, territorial expansion and natural growth. It increases the demand of accommodation of people and due to this demand the house rent has been rising dramatically day by day. About three quarter of Dhaka`s total population live on rented houses. Nearly 65 percent or more of their monthly income is spent on paying the rent. The tenants are in a hostage condition in the hand of landlords. The landlords fix the house rent as their wish. Even it is seen sometimes rent is increased yearly by the landlords four or five times. The general dwellers of this city are highly oppressed by the house owners for a long time and they behave like the tyrant and impose house rents without following any prescribed guidelines. There should be a standard to fix the house rent that can stop this aggressive behavior of the landlords. In this paper we build a neural network that can estimate the median price of a home in a neighborhood described by forty demographic

attributes. Neural networks are very good at function fit problems. A neural network with enough elements (called neurons) can fit any data with arbitrary accuracy. They are particularly well suited for addressing non-linear problems. Given the non-linear nature of real world phenomena, like house valuation, neural networks are a good candidate for solving the problem. The thirteen neighborhood attributes will act as inputs to a neural network, and the median home price will be the target. The network will be designed by using the attributes of neighborhoods whose median house value is already known to train it to produce the target valuations.

2. Background Study

A feed forward neural network is a biologically inspired classification algorithm. It consists of a (possibly large) number of simple neuron-like processing units, organized in layers. Every unit in a layer is connected with all the units in the previous layer. These connections are not all equal; each connection may have a different strength or weight. The weights on these connections encode the knowledge of a network. Often the units in a neural network are also called nodes. Data enters at the inputs and passes through the network, layer by layer, until it arrives at the outputs. During normal operation, that is when it acts as a classifier, there is no feedback between layers. This is why they are called feed forward neural networks. In the following figure we see an example of a 2-layered network with, from right to left: an output layer with 2 units, a hidden layer with 4 units, respectively. The network has 3 input units.
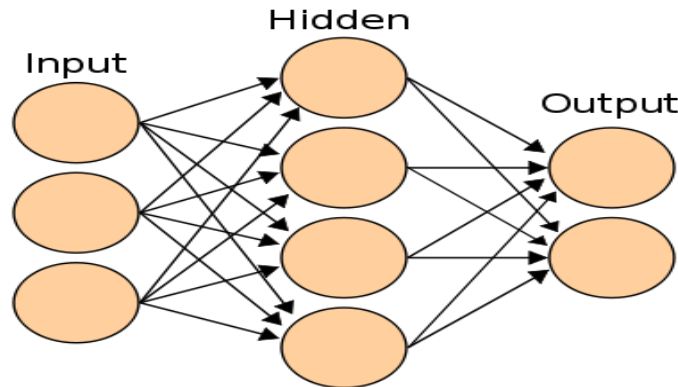

Figure 1. Multi-Layer Percepton Neural Network

The 3 inputs are shown as circles and these do not belong to any layer of the network (although the inputs sometimes are considered as a virtual layer with layer number 0). Any layer that is not an output layer is a hidden layer. This network therefore has 1 hidden layer and 1 output layer. The figure also shows all the connections between the units in different layers. A layer only connects to the previous layer.

3. Literature Review

The analysis of property data over time has always been fraught with difficulty. Invariably the market evidence (from transactions) involves the sale of different properties for each time period. This has always made analysis difficult because the sample of properties varies for each time period. Early analysis was based on simple statistics such as the mean or median.

However this was improved with the consideration of methods to adjust for the variations in quality. Works by Bailey *et al.*, (1963), Goodman (1978), Palmquist (1980), Nourse (1982) and Hendershott & Thibodeau (1990) helped to develop a body of methodology to consider constant quality indexes. This came to a head at the start of the 1990's with a paper by Mankiw & Weil.(1989). In this paper the authors challenged estimates on future housing estimates by using constant quality methodologies. This brought about a flood of papers in response. Works such as those by Case et.al.(1991) Engelhardt & Poterba. (1991), Hamilton (1991), Hendershott (1991), Holland (1991) Meese & Wallace (1991).Woodward(1991) helped to develop the methodological arguments. Various studies have investigated the use of constant quality price index methodologies in the Australasian region. Research by Rossini, et al (1995),Costello & Schwann (1996) Rossini (1996a), Rossini (1996b) and Schwann (1997) have all contributed to the regional knowledge.The opportunity for the use of Artificial Neural Networks (ANN) to estimate price indices has been investigated in recent years. For example Borst (1991) reported the use of ANN to data sets of family residences in New England. Tayand Ho (1992, 1994) examined sets in Singapore using 833 residential apartment properties for training and tested this against 222 case set of similar apartment properties. Do and Grudnitiski (1992) used data from a multiple listing service in California while Evans (1993) worked with residential housing in the United Kingdom. The most recent work comes from Worzala (1995), Borst (1995, 1996), McCluskey (1996a,1996b) and Rossini (1997a, 1997b, 1997c). Rossini's research was based on data from South Australia and demonstrated that the results from artificial neural networks could potentially produce superior results to more traditional econometric models in certain circumstances.

4. Proposed Methodology

Artificial Neural Network (ANN) models were introduced by McCulloch and Pitts (1943) as an alternative to algorithmic programming with some previous work by Karl Lashley in the twenties. In Rumelhart and McClelland (1986), an ANN is presented as a set of operators or neurons, with a small amount of storage capacity, connected numerically one-way by links called axons. Nodes operate with local data supplied through the axons, weighted with some parameters, w_{ij}, linking neurons i and j. A propagation rule establishes how inputs to a neuron are valued and processed. The basic model has an input layer, one or several hidden layers formed by non-observable variables, and an output layer containing the dependent variables. Like any model, the specification is the core task when using ANN: it is necessary to define the network topology (number of hidden layers, and the neurons in each of them), the propagation rule, the transformation of explanatory input variables, and so on. The activation function and the learning rule should also be specified. An excessive number of neurons can originate a lack of forecasting power, due to over parameterization. Computer time to estimate the ANN parameters, learning process, is becoming less relevant with the evolution of the speed in the equipment. There are several learning procedures to estimate the parameters; a widely used technique is the back-propagation method. An ANN is like a non-linear regression or a multivariate regression model, with non observable linking variables. Once the topology and the parameters of the network are specified, it can be presented as an ordinary statistical or econometric model. Neural networks are used with different purposes, such as the estimation of models, classification, forecasting, and so on. Here it is used as a modeling tool and as a practical alternative to the well-known econometric hedonic models. Also, it aims at forecasting the value of properties, so it can be used to measure the downturn occurred in the last few years in the real estate market. Here we use a Multi-Layer Perception (MLP) network with one hidden layer, an input layer with the given thirteen attributes used in

estimating the house rent, and an output layer with the real prices. By considering the following forty factors this paper built the required neural network.

1. Per capita crime rate per town
2. Proportion of residential land zoned.
3. Proportion of non-retail business acres per town.
4. 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 major employment centers.
9. Index of accessibility to radial highways.
10. Full value property tax rate per 10,000 BDT.
11. Pupil-teacher ratio by town.
13. Percent lower status of the population.
14. Type of House
     -Detached
     -Semi-detached
     -Basement
     -Apartment
     -Shanty house
15. Age of the Building
     -0-5 (base category)
     -5-10
     - 10-15
     - 15-20
     -20+
16. Bathroom floor
     -Floor tile
     - Mosaic
     -Normal
17. Toilet
18. Water system
19. Hot water
20. Readymade Furniture
21. Elevator /lift
 22. Garage
 23. Natural gas
24. Distance of market
25. Availability of transport
26. Availability for other facilities of daily life
27. Distance of educational institution
28. Distance of hospital and clinic
29. The environment of neighborhoods
30. Security System
31. Decoration of house
32. Decoration of kitchen
33. Number of balcony

34. Availability of air and light
35. Space of room
36. Distance of police station
37. Distance of recreation centre
38. Distance of working place
39. Distance of religious centre
40. Distance of Bank

5. Experiment and Results

The data set description:

The data used to carry out this research is mainly based on a field survey. A random sample survey was conducted covering 100 flats in each area. For this purpose I actually divided the Dhaka city into five parts namely Mirpur, Dhanmondi, Gulshan Uttara and Old Dhaka. Here the outer east portion of Dhaka is not considered. We considered the rent of these flats as the inputs of our proposed Neural Network. The size of these flats saturates between 750 sq.ft to 3000 sq.ft. The information is presented in tables through frequency distribution. Though i have mentioned forty factors behind the estimation of house rent but in this paper I am working with size of flats behind the estimation of house rent.The main findings of this research study include the rent level of the flats basis upon the range of sizes of the flats. The level of house rent depends on various factors which include house type, location, size of house, utility services, environmental characteristics of the area *etc*.

Table 1. Rent per sq.ft. In Four Areas According to Size of Flats

Size of flats (sq.ft.)	Dhanmondi			Gulshan			Baridhara			Mirpur			Uttara			Old dhaka		
	Freq	%	Per sq.ft. rent (Tk.)	Freq	%	Per sq.ft. rent (Tk.)	Freq	%	Per sq.ft. rent (Tk.)	Freq	%	Per sq.ft. rent (Tk.)	Freq	%	Per sq.ft. rent (Tk.)	Freq	%	Per sq.ft. rent (Tk.)
<=1200	-	-	-	-	-	-	-	-	-	30	30	10.2	-	-	-	20	20	11.2
<=1500	-	-	-	-	-	-	20	20	25.5	30	30	17.6	-	-	-	20	20	17.0
1501-2000	40	40	26.2	-	-	-	20	20	29.3	10	10	22.1	25	25	28.4	20	20	23
2001-2500	20	20	30.3	40	40	32	30	30	34.21	10	10	25.3	25	25	33.4	15	15	26
2501-3000	30	30	34	30	30	36.5	20	20	38	10	10	30.4	25	25	38.5	10	10	32
Above 3000	10	10	38.1	30	30	40.3	10	10	42.63	10	10	36.4	25	25	40.6	5	5	35
Total	100	100		100	100		100	100		100	100		100	100		100	100	
Avg. per sq.ft rent (Tk.)	32.15			36.26			33.93			23.67			35.225			24.033		

Figure 2. Map of Dhaka

Processing of Data:

Data for function fitting problems are set up for a neural network by organizing the data into two matrices, the input matrix X and the target matrix T. Each ith column of the input matrix will have thirteen elements representing a neighborhood whose median house value is already known. Each corresponding column of the target matrix will have one element, representing the median house price.

[x,t] = house_dataset;

size(x)

size(t)

We can view the sizes of inputs X and targets T.Note that both X and T has 506 columns. These represent 506 neighborhood attributes (inputs) and associated median house values (targets).

Input matrix X has thirteen rows, for the thirteen attributes. Target matrix T has only one row, as for each example we only have one desired output, the median house value.

Implementation & Result Analysis:

Here we implement the proposed technique by building a neural network through Matlab software. We use the filtering tool to go further. There need six panels to get the final results. Since the neural network starts with random initial weights, the results of this example will differ slightly every time it is run. The random seed is set to avoid this randomness.

setdemorandstream(491218382)

The input and output have sizes of 0 because the network has not yet been configured to match our input and target data. This will happen when the network is trained.
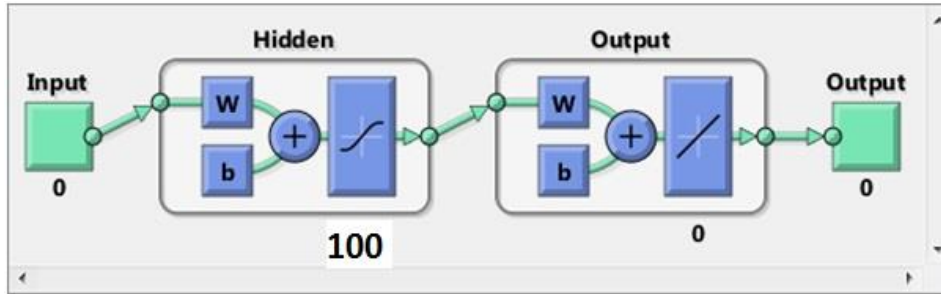
net = fitnet(100);

view(net)

Figure 3. Adapted from the Graphical View of Actual MATLAB Software

First Panel:

A two layer feed forward network with sigmoid hidden neurons and linear output neurons can fit multi-dimensional mapping problems arbitrarily well given consistent data and enough neurons in its hidden layer. The network will be trained with levenberg-Marquardt back propagation algorithm unless there is not enough memory in which case scale conjugate gradient back propagation will be used.

Second Panel:

Here we need to select data set and house pricing. There are thirteen attributes by which we would like to create a neural network which not only estimates the known targets given known inputs, but can generalize to accurately estimate outputs for inputs that were not used to design the solution.

House target= a1*506 matrix of median values of owner occupied homes in each neighborhood in 8000 of taka.

Third Panel:

This part is for validation and testing data. It consists of three attributes.

1. Training: These are presented to the network during training and the network is adjusted according to its error.

2. Validation: These are used to measure network generation, and to halt training when generation stops improving.

3. Testing: These have no effect on training so provide an independent measurement or network performance during and after training.

Now the network is ready to be trained. The samples are automatically divided into training, validation and test sets. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy.

The NN Training Tool shows the network being trained and the algorithms used to train it. It also displays the training state during training and the criteria which stopped training will be highlighted in green.

The buttons at the bottom open useful plots which can be opened during and after training. Links next to the algorithm names and plot buttons open documentation on those subjects.
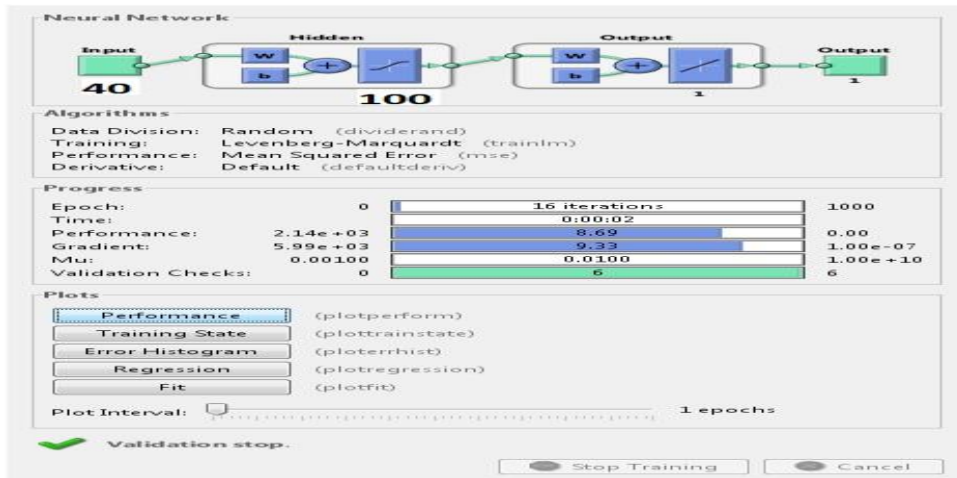[net,tr] = train(net,x,t);

Nntraintool

Network Architecture*:*

Figure 4. Network Architecture

Train Network:

To see how the network's performance improved during training, either click the "Performance" button in the training tool, or call PLOTPERFORM. Performance is measured in terms of mean squared error, and shown in log scale. It rapidly decreased as the network was trained. Performance is shown for each of the training, validation and test sets. The version of the network that did best on the validation set is was after training.
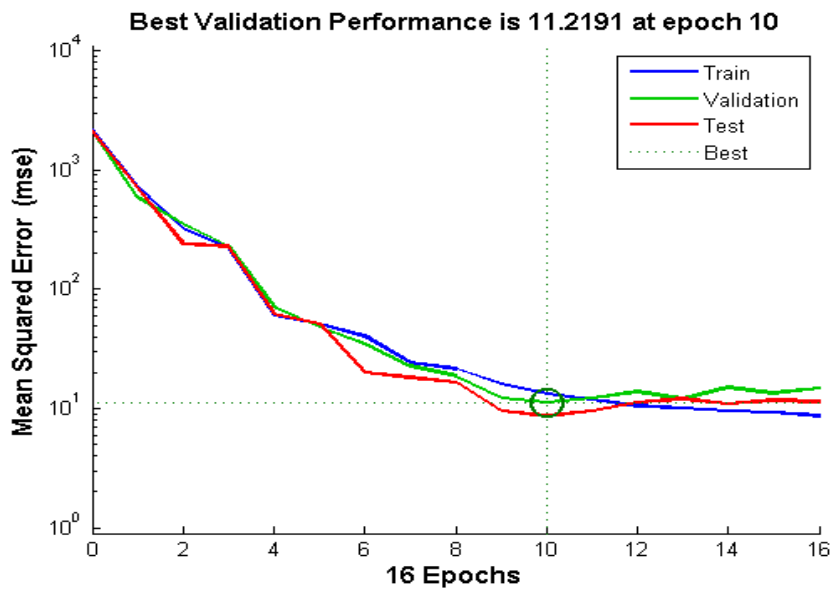
plotperform(tr)

Figure 5. Training Process of Neural Network

Testing the Neural Network:

The mean squared error of the trained neural network can now be measured with respect to the testing samples. This will give us a sense of how well the network will do when applied to data from the real world.

testX = x(:,tr.testInd);

testT = t(:,tr.testInd);

testY = net(testX);

perf = mse(net,testT,testY)

Another measure of how well the neural network has fit the data is the regression plot. Here the regression is plotted across all samples. The regression plot shows the actual network outputs plotted in terms of the associated target values. If the network has learned to fit the data well, the linear fit to this output-target relationship should closely intersect the bottom-left and top-right corners of the plot. The regression line shows that about 92.5% variation of the house rent is explained by the size of the house and indicates a good fit of the data.

If this is not the case then further training, or training a network with more hidden neurons, would be advisable.
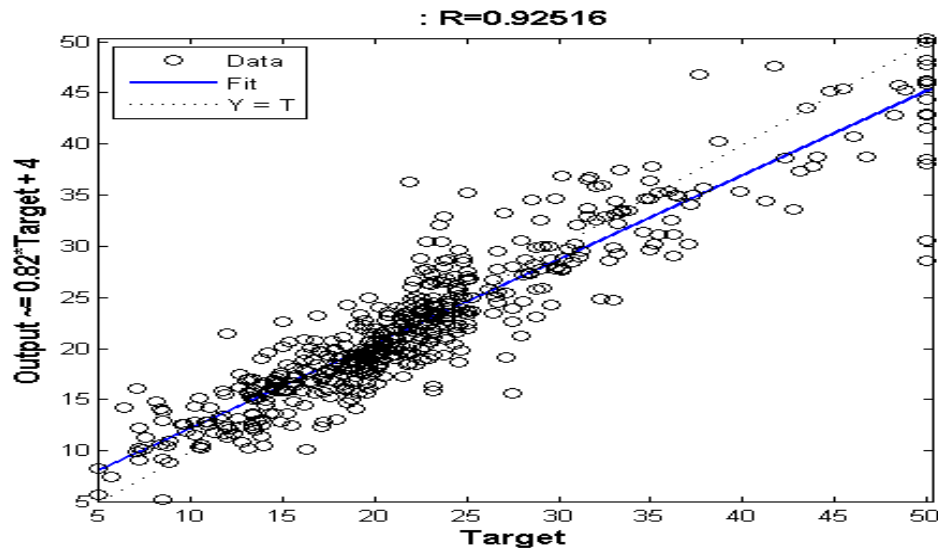
y = net(x);

plotregression(t,y)

Figure 6. Plot of Regression

Another third measure of how well the neural network has fit data is the error histogram. This shows how the error sizes are distributed. Typically most errors are near zero, with very few errors far from that.
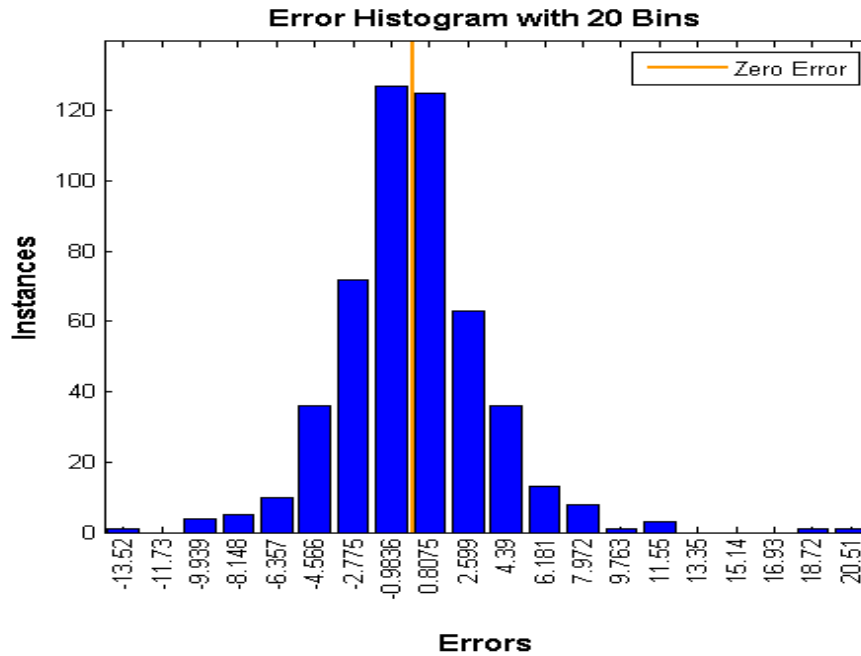
e = t - y;

ploterrhist (e)

Figure 7. Error Histogram

This method illustrated how to design a neural network that estimates the median house value from neighborhood characteristics.

Final Panel: Here we check the result first, if we are not satisfied with our solution we have to return to the architecture and change it to get the desired output. After getting the desired output we should save the results.

% Define inputs & targets

Inputs=houseInputs;Targets=houseTargets;

%Create a fitting network

Hidden layer size=10;

Net=fitnet (hiddenLayerSize);

% setup division of data for training,validation,testing

Net.divideParm.trainRatio=70/100;

Net.divideParm.valRatio=15/100;

Net.divideParm.testRatio=15/100;

%train the network

(net,tr)=train(net,inputs,targets);

%Test the network

Outputs=net(inputs);

Errors=subtract(targets,outputs);

Performance=perform(net,target,outputs);

%view the network

View (net);

Command Window:

>>nnstart

>>whos

Table 2. Output

Name	Size	Bytes	Class	Attributes
Error	1*506	4048	Double	
houseInputs	13*506	52624	Double	
houseTargets	1*506	4048	Double	
Info	1*1	6208	Struct	
Net	1*1	63225	Network	
Output	1*506	4048	Double	

6. Discussion and Limitations

The house rent in Dhaka city has been increasing irrationally in recent times. Most middle class families reside in rented houses in Dhaka. They are struggling to cope with the unbridled rise in house rents. Only about 20-25% of the total inhabitants of Dhaka have their own houses. The remaining 75-80% people live in rented houses. As the demand for rented houses is increasing with the city's rising population, the owners are taking advantage of the situation to raise their rents. In some cases, landlords are also pressuring tenants to leave the house in case the latter are unable to meet the demand. The authorities must settle a standard for fixing the rent of a house. It reduces the injustice over the tenants. And here I propose a design of a neural network that can be used to estimate the house rent. By imposing this method in estimating the house rent the house rent can be fixed upon the given thirteen attributes, so the landlords can't be able to determinate the house rent willingly. Hence there exist some limitations in this technique.

Firstly, it works with only thirteen attributes, but many other factors can be calculated during fixing the house rent. Secondly, when the neural network parameters vary, several system responses can be obtained. By varying these parameters, different results can be obtained when using a neural network structure. In fact, they depend on the numbers of layers, hidden numbers of nodes, the activation function, the initial weights, the bias, the number of patterns and learning algorithm. These characteristics sometimes make complication to get the desired output.

Thirdly, this paper considered only the current year's information of the house rent. The time effect of the house rent, which could potentially impact the estimated results was ignored (the same house rent have different price in different years, assuming that age factor is constant). Finally, the house rent could be affected by some other economic factors (such as exchange rate and interest rate) are not included in the estimation.

7. Conclusion

This paper illustrates the fitting problem, where inputs are matched up to associated target outputs, and we would like to create a neural network which not only estimates the known targets given know inputs, but can generalize to accurately estimate outputs for inputs that were not used to design the solution. This network is designed in purpose of getting a standard for house rent of Dhaka city that the house owners can't be able to raise the house rents according to their whim. To handle the house rent problem a rent controller should be appointed for each ward, that maximum rents for particular areas should be gazette, and that all rents should be paid through banks. It also should be tried to build the neural network that will work more efficiently than the present one.

References

[1] N. Ahmed and N. Khan, "Evolution of House form in Dhaka City, Global Built Environmental Review (GBER), vol. 3, no. 3, (**2001**), pp. 38-48, BBS.

[2] R. Hafiz, "Comfort and Quality of Indoor and Outdoor Spaces of Dhaka: An Analysis of Urban Planning and Design", Global Built Environmental Review (GBER), vol. 4, no. 2, (**2004**), pp. 61-70.

[3] M. Hoek-Smit, "Housing Finance in Bangladesh – Improving Access to Housing Finance by Middle and Lower Income Groups Report Prepared for the Government of Bangladesh", Ministry of Local Government, (**2004**).

[4] Rural Development and Co-operatives, GOB/ UNDP/ UNCHS (Habitat). Islam, N. The Poor's Access to Land and Housing in Dhaka Paper written for the World Bank Dhaka Urban Poverty Study, (**2004**).

[5] A. Adair, S. Mcgreal, A. Smyth, J. Cooper and T. Ryley, "House price and accessibility: The testing of relationships within the Belfast urban area", Housing Studies, vol. 15, no. 5, (**2000**), pp. 699-716.

[6] H. X. H. Bao and A. T. K. Wan, "On the use of spline smoothing in estimating hedonic housing price models: empirical evidence using Hong Kong data", Real Estate Economics. vol. 32, issue 3, (**2004**), pp. 487-507.

[7] O. Bin, "A prediction comparison of housing sales prices by parametric versus semi-parametric regressions", Journal of Housing Economics, vol. 13, (**2004**), pp. 68-84 ss.

[8] A. S. M. Mahbubun Nabi, Md. Kamruzzaman,W. Khalil and F. Khandokar, "Apartment Housing in High Class Residential Areas of Dhaka City:A Case Study of Dhanmondi", Gulshan and Baridhara, (**2004**).

Authors

Samsuddin Ahmed, has been lecturing in CSE since mid of 2010. He is in Computer Science and Engineering from University of Chittagong with highest CGPA till date. His under-grade Thesis was on "Handling Uncertainties in Spatial Feature Extraction". His hobbies include thinking about underlying mathematical formulations in natural phenomena. His research interests include data and image mining, Semantic Web, Business Intelligence, Spatial Feature Extraction etc. He is now serving one of the top most private Universities in Bangladesh named Bangladesh University of Business and Technology (BUBT).

Md.Mahbubur Rahman, has been lecturing in CSE since mid of 2011, he received his B.Sc.Engg. in CSE from Patuakhali Science and Technology University in 2011and continuing his M.Sc. Engg. in CSE at Bangladesh University of Engineering and Technology(BUET), Bangladesh. He is now serving one of the top most private Universities in Bangladesh named Bangladesh University of Business and Technology (BUBT). His research interests are Digital Forensics, Secure and Trustworthy Computing, Data mining, Graph theory, Neural Network.

Sabirah Islam, has recently completed her B.Sc in Computer Science & Engineering (CSE) from Bangladesh University of Business & Technology (BUBT), a reputed university of Bangladesh. She has recently involved in research. Her research interests are Neural Network, Data Mining, Artificial Intelligence, 3D Technology, Graphics and Robotics. She is now working with estimation and decision making model of Neural Network.