

Application of Artificial Neural Networks for Real Estate Valuation

Wojciech WILKOWSKI, Tomasz BUDZYŃSKI Poland

Key words: real estate valuation, artificial neural networks, genetic algorithm

SUMMARY

In the process of massive valuation of real estates models of regression, in particular multiple regression, are applied. At the same time new methods of determination of real estate values are searched for, which will allow to ensure the higher accuracy of estimation.

In the recent years, increased interests in artificial neural networks (ANN) may be noticed all over the world; such networks are being developed as a result of research performed in the field of artificial intelligence. It is a highly sophisticated modelling technique, which allows to project functions of high level of complexity. Artificial neural networks may be applied in many fields of human activity, in all cases where solutions concerning prediction, classification or control should be found. Experiments concerning such networks for real estate valuation were started in the nineties of the twentieth century.

The paper presents general rules of operations of neural networks. It also describes possibilities of using artificial neural networks in two basic solutions concerning real estate valuation: selection of real estate features which highly influence their values and determination of real estate values.

In the practical part of the paper results of experiments concerning utilisation of artificial neural networks for valuation of real estates are presented using an example of apartments and non-built-up areas planned for one-family houses, located in the city of Otwock near Warszawa. First, results of experiments concerning those features of real estates which considerably influence their prices, are discussed. Then, accuracy of determination of real estate values using various types of models of artificial neural networks are compared, including the multi-layer perceptron (MLP), networks of radial base functions (RBF) and linear networks. Obtained values of accuracy of real estate valuation are compared with the values of accuracy, which may be obtained using the widely applied multiple regression method.

At the end conclusions resulting from performed experiments are presented.

Application of Artificial Neural Networks for Real Estate Valuation

Wojciech WILKOWSKI, Tomasz BUDZYŃSKI Poland

1. INTRODUCTION

In the process of real estate valuation, and in particular, in the case of mass valuation, where statistical analysis methods are applied, new methods of determination of a real estate value are searched for, which would allow to achieve higher accuracy of results.

Artificial neural networks (ANN) represent one of methods which might be an alternative for the commonly applied method of multiple regression. They may be applied in the process of determination of real estate values, as well as at the stage of selection of those real estate features, which highly influence their prices.

In the recent years, increased interests in artificial neural networks may be noticed all over the world; such networks are being developed as a result of research performed in the field of artificial intelligence. It is a highly sophisticated modelling technique, which allows to project functions of high level of complexity. Artificial neural networks may be applied in many fields of human activity, in all cases where solutions concerning prediction, classification or control should be found. Experiments concerning such networks for real estate valuation were started in the nineties of the twentieth century.

2. ARTIFICIAL NEURAL NETWORKS

ANN are the highly sophisticated modelling technique, which allows project functions of a very high level of complexity (Statsoft Polska Ltd., 2000). Their name originates from a network of brain nervous cells (McCluskey 1996). ANN have been created as a result of multiyear investigations performed in the field of artificial intelligence, which concerned, inter alia, construction of models of the basic structures which occur in a brain (Statsoft Polska Ltd. 2000a).

A simplified neuron model has been applied in the ANN. It has been defined in the following way: Input signals (values) reach the neuron. They are the primary data values, which enter the network from outside or they are intermediate signals which originate from outputs of other networks, being the elements of the network. Every signal is introduced to the neuron through a connection of a specified strength (weight), which corresponds to the synapse effectiveness in a biological neuron. Every neuron has one threshold value, which specifies the intensity of stimulation required for ignition.

Every neuron calculates the weighted total of its entries, and the level of stimulation, which is thus determined, becomes the argument of the activation function, which calculates the output value of the neuron.

Similarly to construction of a human's brain, ANN consist of many processing points (neurons), arranged in layers and of many mutual connections (synapses) between those points.

Many types of neural networks exist. Multi - Layer Perceptrons (MLP) with one or two hidden layers are mostly used. Besides, the following types of ANN may be distinguished:

- Networks of Radial Basic Functions (RBF),
- Probabilistic Neural Networks (PNN),
- General Regression Neural Networks (GRNN),
- Linear Networks,
- Kohonen Networks.

Architecture of a multi-layer perceptron with one hidden layer is presented in Fig.1.

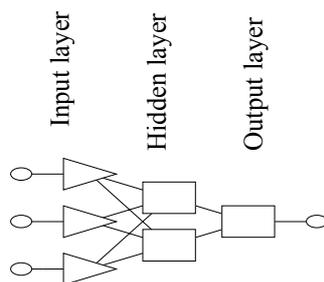


Fig. 1. Architecture of a multi-layer perceptron

Although ANN are computer routines, which simulate operations of neural networks, they are not programmable; they are rather „trainable” on particular examples (McCluskey 1996). In the case of ANN utilisation for real estate valuation such examples are: features of real estates including its selling prices or the level of rent.

Neural networks utilise the parallel structure of data processing (Wiśniewski 1998). Teaching the network is performed by means of input data processing (examples) (McCluskey 1996). Successive passages of data series through the SSN result in fitting of weights of particular connections and threshold values of the neuron in such a way that differences between the result of network operation (the real estate value) and the expected result (the real estate price) are as minimum as possible.

Not all network types and not every number of neurons (including hidden neurons) may give results which satisfy the ANN users. The appropriate type and network architecture should be chosen. The appropriate network is searched by means of training and testing successive variants. In order to accelerate the process some routines, which simulate the ANN operations allow to automate through utilisation of an automatic designer , which tests selected network types of various numbers of neurons (including hidden neurons), in order to find the best network. The best ANN usually determines the network characterised by the smallest validation error (Statsoft Polska Ltd. 2000).

If the user designs the network, he must successively specify the network type, the number of hidden layers, the number of neurons in particular layers, and then train the network.

In the process of ANN designing it is important to make the decision concerning the number of hidden layers and the number of neurons hidden in every layer. Their number depends on complexity of a problem, which should be solved by the neural networks (Wiśniewski 1998). If the number of neurons is too high, it may result in “overtraining” the network; on the other hand, if the number of neurons is too small, the network recognises rules which exist in the data set with insufficient accuracy.

Following (Statsoft Polska Ltd. 2000) the number of all neurons should be selected in such a way that the number of teaching cases is several times higher than the number of weights in the network. In the case of valuation the number of neurons in the input layer may be decreased by means of appropriate selection of features of real estates.

The ANN training consists of three phases: the phase of teaching, the phase of testing and the phase of analysis of results. It requires that the set with examples is divided into two subsets, as the minimum: the teaching subset and the valuating subset. Besides, the test subset may be also created.

Every distinguished subset should contain all examples, which are characteristic for the entire set of data.

In the phase of teaching the network is trained using the teaching subset. It recognises general relations which occur in the data set. In the process of the network teaching the, so-called, teaching algorithms, are applied. For the most popular ANN architecture – the multi-layer perceptron, several teaching algorithms exist. The method of backward propagation of errors is mostly used. Other algorithms include such methods as: the method of coupled gradients, quasi Newton method and Levenberg-Marquardt method.

Then the network operations are controlled in the phase of testing. Tests are performed for the valuating subset, which contains data which do not occur in the teaching subset. This allows for estimation whether the network has been taught the general structure of data or rather it has been adopted to particular values existing in the teaching set. In the first case the error in the teaching set corresponds to the error level in the valuating subset, while in the second case the error level in the teaching subset is considerably smaller than error in the valuating subset. In the course of teaching error values may be controlled in both subsets. If the error in the teaching subset decreases and the error in the valuating subset increases, the training of the network should be terminated, since continuation of teaching will lead to “overtraining”.

If many examples are available, a separate testing subset may be distinguished from the data set. It contains examples which have not been used neither in the teaching process nor in the process of valuating. Testing of the network operations using the discussed subset, in the case

when the small error value is obtained, ensures the user that the network will generate satisfactory results.

The last phase of the network training is analysis of results. In the case of regression issues, as real estate valuation, it comprises specification of the following elements for the above discussed subsets:

- The mean error (being the module of the difference between the assumed value and the obtained output value for the output variable (the real estate value)),
- The quotient of standard deviations for errors and data – being the basic index of the quality of the regression model which has been created by the network.

It should be added that not all networks required the training performed according to the above rules. An example of such a network is the linear network, which has only two layers: the input and the output layers with the linear activation function. This allows to apply methods of determination of weights in the output layer using the conventional least square method. The linear network may become a good reference, which may be used for comparison with ANN network of higher more complexity.

Apart from the above possibilities of using ANN networks in the process of determination of real estate values, ANN may be also useful in the process of selection of real estate features, which highly influence the real estate prices. In order to select an appropriate combination of entries (the real estate features) the following methods may be applied: the genetic algorithm, the stepping forward or the stepping reverse methods.

Genetic algorithms are the technique which allows to look through the high number of combinations (sets of input variables in this case), in order to find the best solution. Utilisation of genetic algorithms is particularly useful when interrelations between optimised variables exist and when impacts of their mutual interactions should be considered.

The stepping forward or the stepping reverse methods allow to find the appropriate set of variables (features) by adding or eliminating successive variables from the data set, respectively.

2. TESTING UTILISATION OF ARTIFICIAL NEURAL NETWORKS IN REAL ESTATE VALUATION

Tests concerning utilisation of ANN networks in real estate valuation have been performed on the examples of the market of non built-up areas, planned for one-family houses and of the market of apartments located in the city of Otwock near Warszawa. Accuracy of specification of real estate values using the ANN networks has been compared with results obtained with the use of the multiple regression method.

2.1. Non Built-up Areas Planned for One-family Houses

Possibilities to apply the ANN for real estate valuation on the example of the market of non built-up areas, planned for one-family houses, have been tested for two variants, diversified

with respect to the number of transactions: variant I - several dozens of transactions, variant II – more than 100 transactions.

2.1.1 Variant I

Data on 59 transactions concerning land, non built-up real estates, which are planned for construction of dwelling houses, performed in the period of 2000-2001 and located in the central Otwock, have been tested. Prices of 1 sq.m. were updated for January 2002.

2.1.1.1 Selection of Features

Those features, which may influence land prices on the local real estate market, being the central part of Otwock, have been specified. They are: parcel size, technical infrastructure, access to public transport, neighbourhood, and management conditions.

In order to select features which should be considered in the course of the ANN development, 3 independent methods were applied: the genetic algorithm, the stepping reverse and the stepping forward method, using the computer software Statistical Neural Networks.

Table 1 below presents results obtained with the use of the above methods of feature selection. The following elements were specified for each method: error for the successive set of features and the set of features (the order of features as above, the considered feature have the value 1, the not considered feature has the value 0). An example row from the column 1 – Genetic algorithm means 0.0493 – the error for the set of features (00011) which includes the following features: neighbourhood, conditions of management.

Table 1. Results obtained for various methods of feature selection.

Genetic algorithm	Stepping reverse	Stepping forward
0.0494 00001	0.0478 11111	0.0511 10000
0.0493 00011		0.0502 10100
0.0492 00100		0.0498 10101
0.0490 00101		0.0495 10111
0.0489 00111		0.0494 11111
0.0488 01111		
0.0485 10001		
0.0484 10011		
0.0484 10100		
0.0481 10101		
0.0480 10111		
0.0479 11111		

Basing on results obtained it has been stated that all above features, i.e.: parcel size, technical infrastructure, access to public transport, neighbourhood and conditions of management, considerably influence the real estate prices.

2.1.1.2 Construction of ANN models

In order to specify a real estate value using the ANN method the computer software Statistica Neural Networks has been applied. The “Automatic Designer” in its advanced version, which allows for testing networks of various architecture and for recording them to the set of networks, was used.

In order to construct the ANN, all transactions were divided into the teaching set – 40 cases and the valuating set – 19 cases. Those sets had similar statistical characteristics – the average value was equal to 86.75 zł/s.m. and 85.21 zł/sq.m., and the standard deviation value 21.54 zł/sq.m. and 21.25 zł/sq.m., respectively.

Models of artificial neural networks were constructed for the above set of features. Several types of networks were applied: the multi-layer perceptron (MLP) with one hidden layer, the network of radial basic functions (RBF), and the linear network. 2 networks were selected for each set of feature out of many designed MLP and RBF networks: the best networks and the network with 2 hidden neurons.

Characteristics of selected MLP and RBF networks, as well as linear network are presented in Table 2.

Table 2. Characteristics of ANN variant I

Type	Hidden neurons	Error of teaching	Valuation error
MLP	21	9.01	10.94
MLP	2	13.67	13.88
RBF	4	16.12	15.01
RBF	2	17.40	19.22
Linear	-	14.04	-

The MLP network turned to be the best network among all tested network types. MLP networks of less complicated architecture (2 hidden neurons) than the best networks, which had more hidden neurons, were used for comparing the SSN model with the model of multiple regression. It was the result of the requirement to meet the condition that several cases (transactions) from the training set are to fall per one connection in the network.

Accuracy of determination of real estate values using the MLP network were compared with accuracies obtained with the use of the multiple regression model. Statistically correct model of multiple regression was constructed with the use of the following features: parcel size, access to public transport, conditions of management (the model constructed with the use of all 5 above features was not correct from the statistical point of view).

The accuracy of the MLP model with 2 hidden neurons was slightly higher than the above model of multiple regression. The average deviation for the MLP model equaled to 13.75 zł/sq.m., and for the multiple regression model – 14.29 zł/sq.m.

2.1.2 Variant II

Data concerning 114 transactions of the period 2000-2001, related to land, not built-up real estates, planned as a location of housing investment, located in Otwock were the subject of analysis. Prices of 1 sq.m. of lands were updated for January 2002.

2.1.2.1. Selection of features for the ANN model

Those features, which may influence the land prices in the city of Otwock, were specified. They are: location, parcel size, technical infrastructure, access to public transport, neighbourhood and conditions of management.

In order to select features which should be considered in the process of construction of the ANN model, 3 independent methods we applied: the genetic algorithm, the stepping reverse and the stepping forward method. Table 3 below presents results obtained with the use of those 3 methods of feature selection. For each method the following elements were specified: error value for the successive set of features and the set of features (the sequence of features as above, the considered feature has the value 1, the feature which is not considered – the value 0).

Table 3. Results obtained for various methods of feature selection

Genetic algorithm	Stepping reverse	Stepping forward
0.0304 000001	0.0303 111111	0.0299 100000
0.0304 000011		0.0265 100100
0.0291 000100		0.0251 101100
0.0289 000101		0.0244 101101
0.0289 000111		0.0244 111101
0.0283 001100		0.0243 111111
0.0282 001101		
0.0281 001111		
0.0281 011101		
0.0281 011111		
0.0272 100000		
0.0270 100001		
0.0270 100011		
0.0258 100100		
0.0256 100101		
0.0255 100111		
0.0250 101100		
0.0248 101101		
0.0248 101111		
0.0247 111111		

As a result of utilisation of those methods it has been stated that all features, i.e.: location, parcel size, technical infrastructure, access to public transport, neighbourhood and conditions of management considerably influence the real estate prices.

2.1.2.2. Construction of ANN Models

In order to construct the ANN the set of transactions was divided into the teaching set – 76 cases and the valuating set – 38 cases. Statistical characteristics of those sets were similar – the average value equalled to 61.23 zł/sq.m. and 59.65 zł/sq.m. and the standard deviation equalled to 32.13 zł/sq.m. and 30.33 zł/sq.m., respectively.

Models of artificial neural networks were constructed for the above set of features. The same types of network which were used for Variant I, were also applied for Variant II. 2 networks were selected for every set of features from all designed MLP and RBF networks: the best network and the network with 3 hidden neurons. Characteristics of selected MLP and RBF networks and the linear network are presented in Table 4.

Table 4 Characteristics of ANN variant II

Type	Hidden neurons	Error of teaching	Valuation error
MLP	5	7.68	8.68
MLP	3	8.52	8.82
RBF	15	8.35	10.14
RBF	3	10.82	12.16
Linear	-	11.09	-

Again, the MLP network turned to be the best network. The MLP networks of less complicated architecture (3 hidden neurons) than the best networks having more hidden neurons, were used for comparison of the ANN model with the multiple regression model. This was influenced by the requirement to meet the condition that several cases (transactions) in the teaching set should fall on one connection in the network.

Values of accuracy of determination of the real estate value with the use of the MLP network were compared with values of accuracy obtained using the multiple regression model. The statistically correct multiple regression model was constructed with the use of the following features: location, parcel size, access to public transport, neighbourhood and conditions of management (the model constructed with utilisation of all 6 features mentioned above was not correct from the statistical points of view).

The accuracy of the MLP model with 3 hidden neurons was considerably higher than in the case of the above discussed multiple regression model. The average deviation for the MLP model equalled to 8.62 zł/sq.m., and for the multiple regression model it equalled to 11.55 zł/sq.m.

2.2. Apartments

Possibilities of using ANN for the needs of real estate valuation on the example of the housing market were investigated basing on 105 transactions related to apartments, which took place in the period 2000-2001, for real estates located in the central part of Otwock. Prices of 1 sq.m. of apartments were updated for January 2002.

2.2.1 Selection of Features for the ANN Model

Those features, which – on the local market of apartments - may influence the price of 1 sq.m. of an apartment located in the central part of Otwock, were specified. They are: the level of wear, standard of an apartments, size of an apartment and location on a particular floor. Due to the lack of possibilities to acquire data concerning the standard of an apartment, this feature was neglected in further analysis, although it considerably influences the price of an apartment. Therefore ANN models were constructed assuming no access to the full information on real estates.

In order to select features which should be considered in the process of construction of the ANN model 3 independent methods were applied: the genetic algorithm, the stepping reverse and the stepping forward method. As a result of utilisation of the genetic algorithm and the reverse stepping method, it has been stated that any feature does not considerably influence the price of an apartment. It is a very particular case. It may be only explained by the low diversification of prices on the local real estate market and by the fact, that the standard of an apartment has not been considered in the process of analysis. For further analysis the following 3 features were assumed: the level of wear of the building, size of an apartment and location on a particular floor.

2.2.2 Construction of ANN Models

In order to construct the ANN model the set of transaction was divided into the teaching set – 70 cases and the valuating set – 35 cases. Those sets have similar statistical characteristics – the average value equalled to 2177.25 zł/sq.m. and 2171.27 zł/sq.m. and the value of standard deviation: 229.34 zł/sq.m. and 225.43 zł/sq.m., respectively. Models of artificial neural networks were constructed for the above set of features. The same types of networks were used as in section 2.1. Two networks, i.e. the best network and the network with 3 hidden neurons, were selected from many designed MLP and RBF networks.

Table 5 presents characteristics of selected MLP and RBF, as well as the linear networks.

Table 5 Characteristics of ANN

Type	Hidden neurons	Error of teaching	Valuation error
MLP	13	214,08	206,47
MLP	3	220,47	210,28
RBF	10	213,00	207,39

RBF	3	222,74	227,26
Linear	-	215,80	-

The MLP network turned to be the best network from tested types of networks. However, it should be noticed, that the RBF and the linear networks turned to be slightly worse.

In order to compare accuracy of determination of real estate values using the MLP network several models of multiple regression were constructed. Unfortunately, none of those models was statistically correct and none of them was used for the needs of comparing.

3. CONCLUSIONS

Artificial neural networks may be applied for real estate valuation. They may be applied both, in the process of determination of real estate values, as well as for selection of real estate features which highly influence the real estate prices.

Artificial neural networks are an alternative solution for the multiple regression method used for real estate valuation.

The accuracy of determination of real estate values using the ANN method is higher than the accuracy obtained with the use of the multiple regression method for real estate markets of high number of transactions – above 100 transactions. In such a case utilisation of the ANN is highly recommended.

On the other hand, for the markets of medium number of transactions – several dozens of transactions – utilisation of the ANN for real estate valuation results in the slightly higher accuracy of estimation of real estate values. Those methods may be interchangeably applied. Out of the analysed ANN models, the highest accuracy of determination of real estate values may be reached with the use of a multi-layer perceptron (MLP).

As it was mentioned, the ANN may be also applied in the process of selection of those real estate features which highly influence their prices. This, among others, is enabled by the genetic algorithm, which allows to look through the high number of combinations (a set of features in the discussed case), in order to find the best solution.

REFERENCES

- McCluskey W.J. 1996. *Zastosowanie sztucznej inteligencji w wycenie masowej dla potrzeb taksacji nieruchomości*. (Utilisation of artificial intelligence for real estate valuation for taxation purposes), Wycena nr 5(28), Olsztyn.
- Statsoft Polska sp. z o.o. 2000, *podręcznik elektroniczny STATISTICA Neural Networks* (Electronic Manual).
- Wiśniewski R. 1998. *Zastosowanie sztucznych sieci neuronowych do wyceny masowej*. (Utilisation of artificial neural networks for mass valuation), Wycena 1(36), Olsztyn.

BIOGRAPHICAL NOTES

Professor Wojciech Wilkowski

Academic experience: Prof. Wojciech Wilkowski Ph.D. D. Sc, Warsaw University of Technology, Faculty of Geodesy and Cartography, Institute of Applied Geodesy

Current position: Head of the Institute

Practical experience: cadastral surveying, land management, real estate valuation

Member of the Polish Society of Surveyors, Member FRICS

Tomasz Budzyński Ph.D.

Born in Warsaw in 1974. Studies of Geodesy and Cartography at the Warsaw University of Technology. Graduated his (M.Sc.) in Geodesy in 1998. Obtained his Ph.D. with a dissertation "Research on an effective method of determination of cadastral value" at the Warsaw University of Technology in 2005. Current position: full-time research worker at the Warsaw University of Technology (Institute of Applied Geodesy of Faculty of Geodesy and Cartography). Member of the Polish Society of Surveyors

CONTACTS

Wojciech Wilkowski

Plac Politechniki 1

PL 00-661 Warsaw,

POLAND

Tel.+ 48 22 625-15-27

+ 48 22 660-73-69

Fax: + 48 22 625-15-27

Private: (48-22) 674-74-11

E-mail: wwilk@gik.pw.edu.pl

Tomasz Budzyński

Plac Politechniki 1

PL 00-661 Warsaw,

POLAND

Tel. + 48 22 625 15 27

+ 48 22 660 75-89

Fax: + 48 22 625 1527

Mobile: + 48 603 196198

E-mail: tbudzynski@o2.pl