Application of a Neural Network Model for Solving Prediction Problems in Business Management

Kosa Golic, *Member, IEDRC, IACSIT*

***Abstract*—In the field of business management many prediction problems such as prediction of market demand, product prices, project cost flow, product quality, etc., are very complex, uncertain, dynamic, mutable, and usually they are not adequately modeled by statistical or classical mathematical methods based on crisp sets and traditional logic. This paper shows benefits that could be achieved by applying a model based on Artificial Neural Network (ANN) for solving certain type of prediction problems, because it can better deal with uncertainty, partial truth, incomplete data and complexity. A model based on ANN for solving prediction problems is first defined and explained. A case study for price estimation of the apartments in the city Budva, Montenegro, is considered and the results are discussed. The main factors influencing the apartment prices are established and analyzed. The benefits of the ANN model are pointed out.**

***Index Terms*—Apartment price calculation, artificial neural network, prediction model, price evaluation.**

I. Introduction

Prediction problems are very frequent class of problems encountered in many fields of management. For example: calculation of product prices, forecasting time of building construction, prediction of market demand, project cost flow, prediction of product quality, etc. These problems can be mathematically defined as the problems of finding the relationship between two sets of variables.

Most of the traditional methods used to solve this kind of problems are based on regression techniques whereby the best fit is sought. In practice, these methods usually are not capable of giving satisfactory solution because the relations between the considered variable (market demand, product price, etc.) and the factors on which those variables depend on are very complex. Apart from this, the factors that determine the shape of relations are often difficult to quantify or it is not clear which factors the considered variable depends on and degree of effect such factors may have. Also, the relationship between the inputs and outputs in the regression models are very complex since there could be some unknown combined effect.

Further, one of the difficulties in applying these models is accounting for the existing correlation among the input variables modeled as random variables, even if the correlation among random variables is known. Thus, it is difficult to perform such multi-attribute nonlinear mapping by using a regression model. It has been proved that problems with multi-attributes can be solved better by ANNs then by traditional methods [1]. Also, it has been proved that any arbitrarily irregular patterns can be mapped by ANNs [2]. Apart from this, ANNs can learn by themselves, generalize solutions, and respond adequately to highly correlated and incomplete data.

Manuscript received December 19, 2012; revised January 25, 2013.

Kosa Golic is with the School of Construction Management, University Union Nikola Tesla, 11000 Belgrade, Serbia (e-mail: kosa.golic@live.com).

II. Definition of the ANN Model

The model for solving prediction problems is defined following the logic shown in Fig. 1. A sample three-layer neural network illustrated in Fig. 2, is used in this model. There are many publications describing development and theory of ANN from the introductory level to more advanced stages. A detail and comprehensive description of several types of ANNs is given in [3], [4]. A clear summary of the feed-forward network and the back-propagation method for network training is presented in [5]. A statistical interpretation of the methods used to train ANN is given in [6]. In the model proposed, a feed-forward ANN with back propagation training algorithm is used. Several phases (steps) are distinguished in order to facilitate problem solving.

A. First Step: Definition of Input and Output Data

Definition of the input and output variables is a delicate part of the problem solving process. It is important to choose only the input variables which are influencing the output in order not to overload the network.

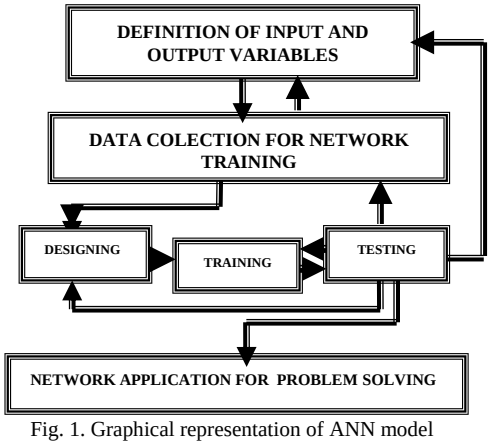

Fig. 1. Graphical representation of ANN model

A consequence of that can be a request for a larger set of training data or a greater number of hidden nodes/layers. In general, output of the network is a vector of variables $Y=(y_1, y_2,...y_n), j=1, 2,...n,$ which are to be calculated (project price, market demand, client satisfaction, product reliability, etc.). Input of the network is a vector of variables $X=(x_1, x_2,...x_m), i=1, 2,...m,$ on which the output vector Y

DOI: 10.7763/JOEBM.2013.V1.32

depends on (quantities of the materials, characteristics of the product, characteristics of the technology production, etc).

Generally, for each project the number of input/output variables, (*m* and *n*) are different. In special cases, the output vector can have only one variable (construction time, market demand or degree of the product quality). It should be pointed out that the input or output variables need not to be mutually independent which gives a great flexibility to this model in solving practical problems.

B. Second Step: Gathering a Set of Data for Network Training

After defining the set of input and output variables, the next step is the collection of a set of training data. That is, for each vector of the input values a corresponding set of the output values should be collected as well. In other words, a set of data for network training consists of the pairs of values of the input-output vectors, $\left(X^k, Y^k\right)$, $k = 1, 2, ...q$, where q is the total number of training vectors. On the basis of these vectors the ANN model learns to approximate required output variables by determining the strengths of connections (weights) between the networks nodes [7], [8].

The number of training vectors to be supplied depends on the total number of input-output variables and complexity of their interrelationships. Generally, the greater the number of variables the greater the training set is required. It is recommended that training data cover whole expected field of input values as evenly as possible.

C. Third Step: Network Design, Training and Testing

Network design directly depends on the number of input-output variables as well as the complexity of their interrelationships. The number of input/output nodes of the network is equal to the number of input/output variables, i.e., to each variable is assigned one node, Fig. 2. Number of the hidden layers and number of their nodes is determined experimentally during the phase of network training and depends on the number of input and output variables as well as the complexity of their interrelationships [4]. Overall, the greater the number of variables the greater is the number of hidden layers and their nodes. For the practical problems, in most cases, two hidden layers with an appropriate number of nodes are usually sufficient for successful modeling and problem resolving.

Satisfactory level of the network training is checked by the testing data (contains both input and output values). When the output values calculated by the model "are close" to the output values from the testing set of data, i.e. the error is within an assigned interval, the training process is finished and the model is ready to be used. Otherwise, the number of the hidden nodes and/or layers should be modified until the satisfactory results are reached.

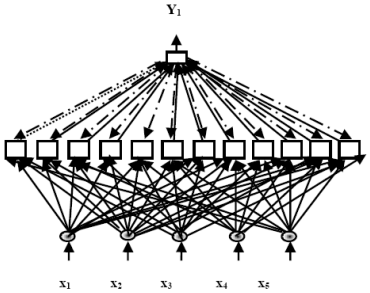

Fig. 2. Graphical representation of ANN for the apartment price calculation.

A degree of required precision for the ANN models is established in advance (before network training) and for each problem can be different.

III. Case Study

A problem of the apartment price estimation in the city Budva, Monte Negro, is considered. Evaluation of the apartment prices is important for all those who have intention to invest money in real estate, construct new flats or sell one, as well as for the banks allowing loans to the clients investing in the real estate sector. There are many factors influencing apartment price. The most important are given in Table I. They are divided into three categories: General characteristics, Interior apartment quality, and Building quality.

In general, evaluation of all factors influencing the apartment price, Table I, is a very demanding and expensive task and therefore is rarely executed. Also, many of the listed characteristics (maintenance of the building, quality of the bath and kitchen furnishing, quality of the façade and roof construction, apartments view, etc.), are subjective categories and could vary significantly depending on the estimators preferences. Therefore, an approximate value as close as possible to the real value is usually sought.

TABLE I: The Factors Influencing Apartments Prices

General characteristics	Apartment Area, City zone, Number of rooms, Apartments floor position , Apartment view, Local surroundings of the apartment, Noise, Proximity of the important buildings (hotels, parks, universities, etc.).
Interior apartment quality	Apartment insolation, Bathroom and kitchen furnishing, Walls and floors finishing, Carpentry.
Building quality	Maintenance of the building, Year of construction, Quality of the façade, Quality of the roof construction, Global quality of the building.

Apart from the listed characteristics (factors), a great influence on the apartments prices has political and economical situation in the country. Also, global economical and political situation is starting to have an increasing impact on the real estate sector of almost each single country in the world. These factors are not considered in this paper but they could be the topic of further research.

Taking into account the availability of the data recorded by the Real Estate Agencies in Budva, Monte Negro, on one side, and the most important characteristics of the apartments regarding the costumers opinion on the other, the following factors are taken into consideration for the apartment price prediction:

1) Apartment area,
2) City zone,
3) Number of bedrooms,
4) Interior apartment quality,
5) Building age.

Total number of the data collected in the Real Estate Agencies in Budva, Monte Negro, was 237. A part of them is shown in Table II. For the assessment of the interior apartment quality, the agencies were used descriptions

"excellent", very good, and "good", but in cases of a poor quality, they left out any description. It meant that the interior apartment quality was poor. In order to quantify these qualitative ratings: excellent, very good, good and poor, the numbers 6, 4, 3 and 1 are ascribed to them respectively. Quantification of the city zones is given as follows: 1 for downtown, 2 and 3 for the zones less attractive. For the "building age", quantification is made as follows: number 1 for the building less than 10 years old, 2 for less than 20 years old building, and so on, up to 7 meaning more than 60 years old building. The values of the apartments area and their number of bedrooms are taken in their original state.

TABLE II: INPUT DATA AND CORRESPONDING RESULTS

Building age	Apartment area (m^2)	Number of bedrooms	City zone	Interior quality	Apartment price (euro)	Calculated apartment prices (euro)
3	**65**	**3**	**2**	**1**	**43500**	**45476**
7	**84.77**	**3**	**2**	**3**	**64000**	**64901**
5	**76**	**3**	**2**	**6**	**77800**	**73870**
2	**120**	**3**	**3**	**4**	**97200**	**98656**
1	**111**	**3**	**3**	**3**	**77000**	**72091**
2	**60**	**3**	**3**	**4**	**51200**	**54550**
3	**72.5**	**3**	**3**	**6**	**70000**	**67017**
1	**41**	**1**	**2**	**6**	**41000**	**39107**
3	**57.7**	**2**	**2**	**3**	**47000**	**47552**
6	**80**	**2**	**3**	**3**	**60000**	**59392**
1	**57**	**2**	**3**	**3**	**38000**	**41240**
1	**90**	**3**	**1**	**4**	**94600**	**89958**
2	**90**	**3**	**1**	**3**	**83500**	**86024**

For the calculation of the apartments prices, the network with 5 nodes in the input layer, 1 in the output layer and 15 in the hidden layer is designed. In the hidden nodes the sigmoid transfer function is used and for the output node the linear one. For the network training the back-propagation algorithm is applied. During network training, the number of nodes in the hidden layer was modified from 15 to 12 in order to decrease learning error. All calculations were done using the computer software for ANNs which is incorporated in the computer package Matlab 6.5.

IV. RESULT ANALYSIS

On the basis of the results obtained for the set of testing data, Table II, it can be concluded that the ANN model is well trained because 85% of the results have deviations less than 5% and the rest of them have maximal deviation no more than 8%. This is a satisfactory solution for the problem in question. A greater accuracy could be obtained by providing a set of data covering more evenly the field of input data for the "interior apartment quality" and "building age".

From the results obtained, it is shown that variables "apartment area" and "city zone" have the greatest impact on the apartment price while the "interior apartment quality" has a moderate one and the "building age" little less of it. Variable "number of bedrooms "has the smallest impact on the apartment price.

These results can be explained by the fact that "apartment area" and "city zone" are the most important criteria (attributes) for the costumers and they are not willing to compromise on them. Conversely holds for the "interior apartment quality". The impact of the "building age" on the apartment prices is complex, because a durability of the building depends also on its maintenance, type of materials used, construction process applied, etc., and therefore inexpert costumers could be easily confused with external appearances of facades. Regarding the low impact of the variable "number of bedrooms", it can be concluded that for the majority of costumers this is the least important factor and that they are willing to pay for apartment adaptation according to their needs if the other factors are satisfactory.

V. CONCLUSION

Application of the ANN model is convenient for solving prediction problems where the multi-attribute nonlinear mapping is required. The input variables need not be mutually independent. This fact is very important because majority of the practical problems are presented in that way. Complexity of the relations between input/output variables is not a limiting factor for the ANN model, which can therefore be applied for solving a large number of practical problems. Apart from this, one of important advantages of this method is its capacity to solve problems even if the set of input data is not complete. This has been demonstrated in the case study regarding the prediction of apartments prices. Namely, for the apartment price estimation seventeen characteristics are proposed to be examined, in Table I. However, for the model in question only five of them were taken into consideration and the results were satisfactory, as shown in Table II. This facilitates the problem solving to a great extent.

Another important fact is the simplicity of the model application. On the software market place there are many computer programs for the ANN calculations which facilitate the model use.

The main disadvantage of the ANN model is in the difficulty of providing input data. However, this limitation in the "computer era" is decreasing as the electronic way of data storage is becoming prevalent.

REFERENCES

[1] T. Masters, *Practical Neural Networks Recipes in C++*, Academic Press London, 1993.
[2] P. R. Hush and B. G. Horne, "Progress in supervised neural networks," *IEEE Signal Processing Journal,* pp. 8-39, January 1993.
[3] B. Widrow, D. Rumelhart, and A. Lehr, "Neural networks: Application in industry, business and science," *Communications of the ACM,* vol. 37**,** no. 3, pp. 93-105, 1994.
[4] T. Kohonen, "An introduction to neural computing," *Neural Networks,* vol. 1, no. 1, pp. 3-16, 1998.
[5] S. Karmarthi, V. Sanvido, and R. Kumara, "Neuroform–neural network system for vertical formwork selection," *ASCE Journal of Computing in Civil Engineering,* vol. 6 , no. 2, pp. 178-199, 1992.

[6] E. Levin, N. Tishby, and S. A. Solla, "A statistical approach to learning and generalization in layered neural networks," *IEEE Proceedings*, vol. 78, no.10, pp. 1568-74, 1990.
[7] P. S. Hush and B. G. Horne, "Progress in supervised neural networks," *IEEE Signal Processing Journal*, pp. 8-39, January 1993.
[8] A. H. Boussabaine, "The use of artificial neural networks in construction management: A review," *Construction Management and Economics*, vol. 14, no. 5, pp. 427-36, 1996.

Kosa Golic was born at Tutnjevac, exJugoslavia. She received PhD and MSc in Civil Engineering, major in construction organization and technology, and BSc, major in structural engineering, from the School of Civil Engineering, University of Belgrade, Serbia, 2008, 1993, and 1985, respectively. She was a Research Assistant at Cornell University, Ithaca, SAD, and Queensland University of Technology, Brisbane, Australia. She was employed at the Institute for Material Testing of Serbia, the largest building research institute in exYugoslavia, as a Research Associate. Then she worked with the Municipality of Vozdovac, Belgrade, Serbia, as a Senior Specialist. She is currently with the University Union Nikola Tesla, Belgrade, Serbia, teaching the courses in Construction Management. Her main areas of research interest are construction management including network planning, multi-criteria decision making, applied fuzzy sets, applied artificial neural networks, integrated network fuzzy systems, and linear and nonlinear programming.

Ass. Prof. Dr Golic is a member of the Union of Engineers and Technicians of Serbia (SITS), IACSIT, and Senior Member of IEDRC. She was a member of Committee for "Award for Innovation" at Civil Engineering Fair in Belgrade. She served as a Member of Committee "Investment in Construction" at Municipality of Vozdovac, Belgrade, for several years. She was also a Student Representative at the "Council of Professors and Students" at the School of Civil Engineering, University of Belgrade.