Data Optimization with Multilayer Perceptron Neural Network and Using New Pattern in Decision Tree Comparatively

[1]Murat Kayri and [2]Omay Cokluk
[1]Department of Computer Science, Yuzuncu Yıl University,
Faculty of Education, Van/Turkey
[2]Department of Measurement and Evaluation, Ankara University,
Faculty of Education, Ankara/Turkey

Abstract: Problem statement: The aim of the present study is to exemplify the use of Artificial Neural Networks (ANN) for parameter prediction. Missing value or unreal approach to some questions in scale is a problem for unbiased findings. To learn a real pattern with ANN provides robust and unbiased parameter estimation. **Approach:** To this end, data was collected from 906 students using "Scale of student views about the expected situations and the current expectations from their families during learning process" for the study entitled "Student views about the expected situations and the current expectations from their families during learning process". In the study, first the initial data set gathered using the measurement tool and the new data set produced by Multi-Layer Receptors algorithm, which was considered as the highest predictive level of ANN for the research were individually analyzed by Chaid analysis and the results of the two analyses were compared. **Results:** The findings showed that as a result of Chaid analysis with the initial data set the variable "education level of mother" had a considerable effect on total score dependent variable, while "education level of father" was the influential variable on the attitude level in the data set predicted by ANN, unlike the previous model. **Conclusion/Recommendations:** The findings of the research show Artificial Neural Networks could be used for parameter estimation in cause-effect based studies. It is also thought the research will contribute to extensive use of advanced statistical methods.

Key words: Artificial neural networks, multi-layer perceptron, Chaid analysis, back-propagation

INTRODUCTION

Artificial Neural Networks (ANN) is an artificial intelligence application developed from the neural (neurological) pattern of human brain's learning spot. First studies on ANN started with modeling the nerve cells forming the brain and the application of these modeling's into the computer systems (Haykin,1999). Afterwards, it has become common in many fields in parallel with the development in computer systems. ANN has efficient usage in many fields such as medicine, industry, biology, electronic systems, optimization and social sciences (Golden, 1996).

Artificial Neural Networks could conduct both linear and non-linear model approaches together and so it can get the correlation between variables on a more valid basis (Erilli et al., 2010). ANN is accepted as a strong method that learns the structures of the current data, establishes a new relations network in the real world and conducts many statistical processes such as making parameter estimation, classification, optimization and time series in this relations network in a determined way (Badr *et al*., 2003; Elmas, 2003; Fausett, 1994; Uzun and Erdem, 2005). This method analyzes the data set in three stages. At the first stage, a considerable portion of the data set is used for "training process". ANN tries to detect the relationships between the variables of the data set and so it tries to determine the characteristic of the research pattern. At the second stage, based on the learnings of the first stage, it tries to perceive the model and this process is named the "perceptron process/hidden process". In the perceptron process, the ideal functions that belong to the model are produced and Weights (W_i) of explanatory variable(s) upon the dependent variables are obtained. The third stage is the new model estimation the ANN produces for the real world and this process is named the "output process" (Manel *et al*., 1999).

The first artificial neural network model was developed in 1943 by Warren McCulloch, a neurologist

Corresponding Author: Murat Kayri, Department of Computer Science, Yuzuncu Yıl University, Faculty of Education, Van/Turkey.

and Walter Pitts, a mathematician. The first artificial neural network developed by McCulloch and Pitts was based on a simple electric circuit and the developed model is shown in Fig. 1.

In Fig. 1, "x" represents the predictive variable set in the model, w represents the weight of each predictive variable and vector v represents total weight of independent variables. Finally, y represents the output by artificial neural network. In the network, vector v is obtained by the sum of weighted effects of all predictive variables, since their effects in the model are associated with neurons. Mathematical representation of vector v is shown in Eq. 1:

$$v = \begin{bmatrix} w_1 w_2 w_3 ....w_m w_{m+1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ . \\ x_m \end{bmatrix} \quad (1)$$

$$v = w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_m x_m + w_{m+1} 1$$

ANN includes several sub-types which are used to analyze different models. Gardner and Dorling (1998) created taxonomy for ANN, which is briefly shown in Fig. 2.

ANN tries to learn current data structure within the frame of some learning algorithms. Learning algorithms with using some ANN type discussed in the literature are "Single Layer Perceptron-SLP", "Additive Linear Element-ALE", "Multi-Layer Perceptron-MLP", "Perceptron Neural Network-PNN", "General Regression Neural Network-GRNN" and "Radial Basis Function Networks-RBFN" methods (Gardner and Dorling, 1998; Wieland and Mirschel, 2008). According to the data set being linear or not, ANN not only suggests the ideal learning algorithm, but also shows possibilities of correct definitions for model by different learning algorithms.

Feed forward-back-propagation algorithm and MLP: MLP is one type of neural networks. MLP uses Back-Propagation Algorithm. Therefore, it could estimate the model with least error rate and the ideal weights of independent variables. In other words, it could minimize the difference between the expected
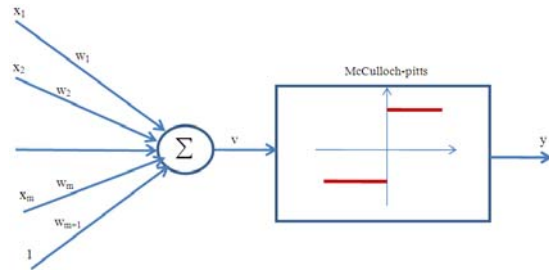

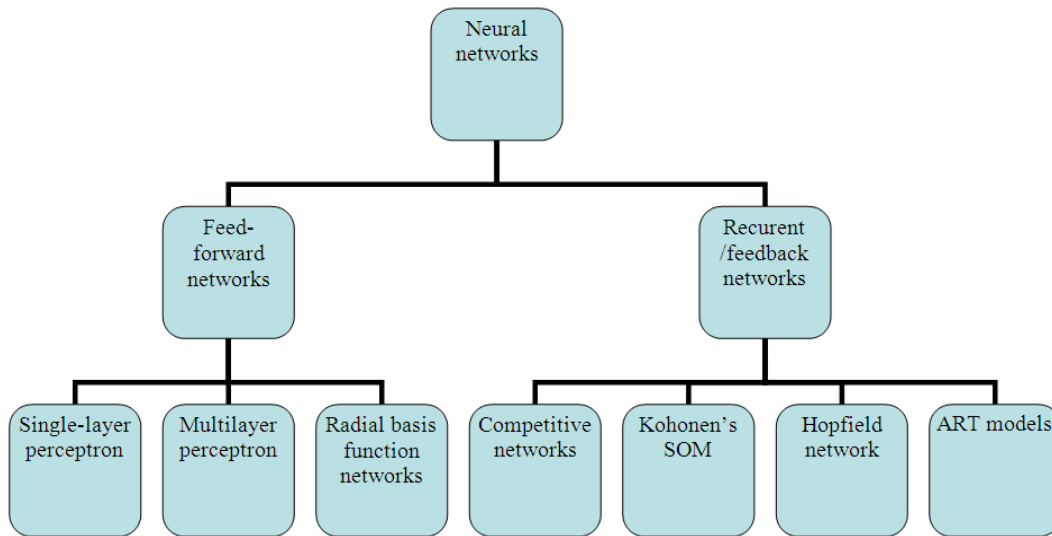Fig. 1: ANN model developed by McCulloch and Pitts (Ayhan, 2009)

Fig. 2: A taxonomy of neural network architectures

and the observed values. In the literature, MLP have been shown to be effective alternatives to more traditional statistical techniques (Gardner and Dorling, 1998; Schalkoff, 1992). Unlike other statistical techniques, MLP does not prioritize assumptions about the distribution of the data set. It is pretty effective in modeling high non-linear functions.

Back Propagation Networks (BPN) is a network which is frequently used. The standard back-propagation algorithms are conjugate gradient algorithms in which net weights move in negative gradient of performance function (Hamzacebi and Kutay, 2004). Back Propagation Networks (BPN), which has many types, is based on standard optimization techniques such as conjugate gradient method and Newton method. Back Propagation algorithm is the process of step-by-step calculation of net weights to minimize network error (Kurt and Ture, 2005). It is the most applied supervised learning algorithm.

Feed forward networks allow a one-way movement from the input to the output, which means there is no feed forward. A typical feed forward ANN consists of the input layer, generally one or two inter-layers (covert layers) and the output layer. Each layer has a varying number of neurons according to the problem under question (Elmas, 2003; Uzun and Erdem, 2005). In Fig. 1, a single covert layer feed forward ANN is shown. There are n number of neurons in the input layer, p number of neurons in the covert layer and m number of neurons in the output layer. Network training is performed through weight adjustment of neural connections in each layer. Weight adjustment procedure is implemented by minimizing error function:

$$E = \frac{1}{2}\sum_{k=1}^{m}\left(y_k - t_k\right)^2 \qquad (2)$$

In error function, y_k represents the output produced by network and t_k represents the expected output value. 1/2 is a constant coefficient which is added for smoothing derivative of the function. The term back propagation algorithm originates from backward weight adjustment to minimize error in the output layer (Hamzaçebi and Kutay, 2004).

Weight change in gradation from tth step of back propagation algorithm to (t+1) th step is shown by the following equation:

$$w_{kj}(t+1) = w_{kj}(t) + \eta\delta k(t)y_j(t) \qquad (3)$$

Here, η is a positive number which is defined as learning rate parameter of back propagation algorithm

and $\delta k(t)$ is the local error of neuron k in the tth step or the difference between the expected output and the observed output of neuron k. $w_{kj}(t)$ represents weight to which neuron k is linked to the input variable in the tth step and weight to which neuron j is linked in the previous step and $y_j(t)$, shows the outputs of neuron j (Kurt and Ture, 2005).

MATERIALS AND METHODS

The research is a correlational study. Correlational survey models are research models which are used to determine the presence and/or the rate of co-change between two or more variables (Karasar, 1991).

Study group-data gathering tools: As it was mentioned before, main aim of the research was to show the use of ANN for parameter estimation over a sample data set. Hence, actual data was used as a result of the study entitled "Views of Students about Current Expectations from Parents in Learning Process and about the Expected Case", where "Questionnaire of Views of Students about Current Expectations from Parents in Learning Process and about the Expected Case" developed by Yılmaz (2009) was used. The measurement tool developed for the research was scaled with five-item Likert type rating. Total explained variance by the measurement tool which consisted of 23 items under a single factor was 35.21. Cronbach Alpha internal consistency coefficient of the measurement tool was 0.91.

The study group consisted of total 906 students; 420 female and 486 male. In the study group, 379 sixth graders, 318 seventh graders and 209 eighth graders were included.

Procedure-data analysis: Kolmogorov-Smirnov test was used to test the normality assumption of score distribution obtained by "Questionnaire of Views of Students about Current Expectations from Parents in Learning Process and about the Expected Case". It was observed that the scale scores were not normally distributed ($p<0.05$).

The predictive variables of the research were as follows: "gender", "grade", "educational background of mother", "educational background of father", "occupation of mother", "occupation of father", "the number of siblings" and "economic income level". Total score from the measurement tool was defined as the predicted variable. The total score mean was 81.34 ± 18.34.

In the constructed regression model, the predicted variable was continuous and all the predictive variables

were nominal (categorical). As a result, testing the initial data set and the new data set by ANN was not possible by parametric regression methods. It is known that semi-parametric methods make strong estimations in the analysis of models where the dependent variable is continuous or categorical and the independent variables are mixed type (categorical, sequencing, interval) (Kayri and Boysan, 2007). Hence, the data set to be comparatively examined by ANN was analyzed by Chi-squared Automatic Interaction Detection method. Chaid analysis is a mixed statistical approach where (tree shaped) sample categorization method and regression analysis are collectively applied.

The research procedures were as follows: Applying Chaid analysis for the data set obtained by the measurement tool application, re-estimating the data set by ANN, re-applying Chaid analysis for the new data set produced by ANN and comparatively examining the results of both analyses.

The data set in ANN application process was individually tested by the following ANN types: Linear, MLP, Radial Based Function-RBF and Generalized Regression Neural Network-GRNN. The results of the learning model calculating "training", "performance-selection" and "output-test" procedures with the highest probability were taken into account during the new data set production. In ANN application, 80% of the data set was used in training and performance-selection and the remaining 20% was used in testing.

RESULTS

As data on the distributions of the variables of gender and grade, two of the predictive variables in the model, was presented under "Study Group", it will not be repeated here. Distributions of the other predictive variables in the model, not presented before, are given in Table 1.

As it is clear from Table 1, almost half of the students' families (47.50%) were in the range of income level from 500 Turkish Liras (TL) and below, most of the mothers (44.26%) and fathers (36.75%) were primary school graduates, most of the mothers were housewives (98.00%), while most of the fathers were workers (20.50%) and self-employed (21.70%)and the number of siblings mostly ranged from 4-6 (37.10%) to 7-9 (42.40%).

Following descriptive statistics procedures of the predictive variables, the data set was examined by ANN for re-estimation. ANN used 444 individuals in the data set for "training", 221 of them for "performance-selection" and the remaining 221 of them for "output-test" process. In Table 2, "accurate estimation" probabilities for the new data set given by the learning algorithms used in ANN application are presented.

Table 1: Descriptive statistics of independent variables belong to students

Variable	Level	f	(%)
Family income	1. Less than 500 TL	431	47.50
(Turkish Liras: TL)	2. 501 TL-1000 TL	224	24.80
	3. Upper 1001 TL	251	27.70
	Total	906	100.00
Mother education level	1. Unliterated	288	31.79
	2. Primary school	401	44.26
	3. Elemantary	114	12.58
	4. High school	75	8.28
	5. University	28	3.09
	Total	906	100.00
Father education level	1. Unliterated	71	7.84
	2. Primary school	333	36.75
	3. Elemantary	186	20.53
	4. High school	205	22.63
	5. University	111	12.25
	Total	906	100.00
Mother occupation	1. Housewife	888	98.00
	2. Officer	10	1.10
	3. Retired	2	0.20
	4. Worker	6	0.70
	Total	906	100.00
Father occupation	1. Unemployed	103	11.60
	2. Worker	182	20.50
	3. Free lance	192	21.70
	4. Marchandise	160	18.10
	5. Retired	74	8.40
	6. Officer	175	19.80
	Total	886	100.00
Siblings number	1. 1-3 siblings	18	2.00
	2. 4-6 siblings	336	37.10
	3. 7-9 siblings	384	42.40
	4. Upper 10 siblings	168	18.50
	Total	906	100.00

Table 2: Summary of active networks

ANN types	Training	Perception performance	Test performance
Linear	0.933	0.991	0.967
RBF	0.924	0.990	0.953
GRNN	0.612	0.939	0.970
MLP	0.965	0.992	0.976

As it is clear from Table 2, MLP type of ANN had the highest estimation level. Therefore, as mentioned in the literature, MLP optimized the data set with least error, taking y = f(x) model into consideration.

The initial data set obtained by the measurement tool and the new data set obtained by ANN were individually examined by Chaid analysis. Tree diagrams obtained by Chaid analysis showed the effects of the predictive variables on the predicted variable and the significance level of the effects. In Fig. 3, the diagrams of the initial data set and in Fig. 4, the diagrams of the new data set obtained by ANN are presented.

As it is clear from Fig. 3, "educational background of mother" had an important effect on the dependent variable of total score. It was observed that the income
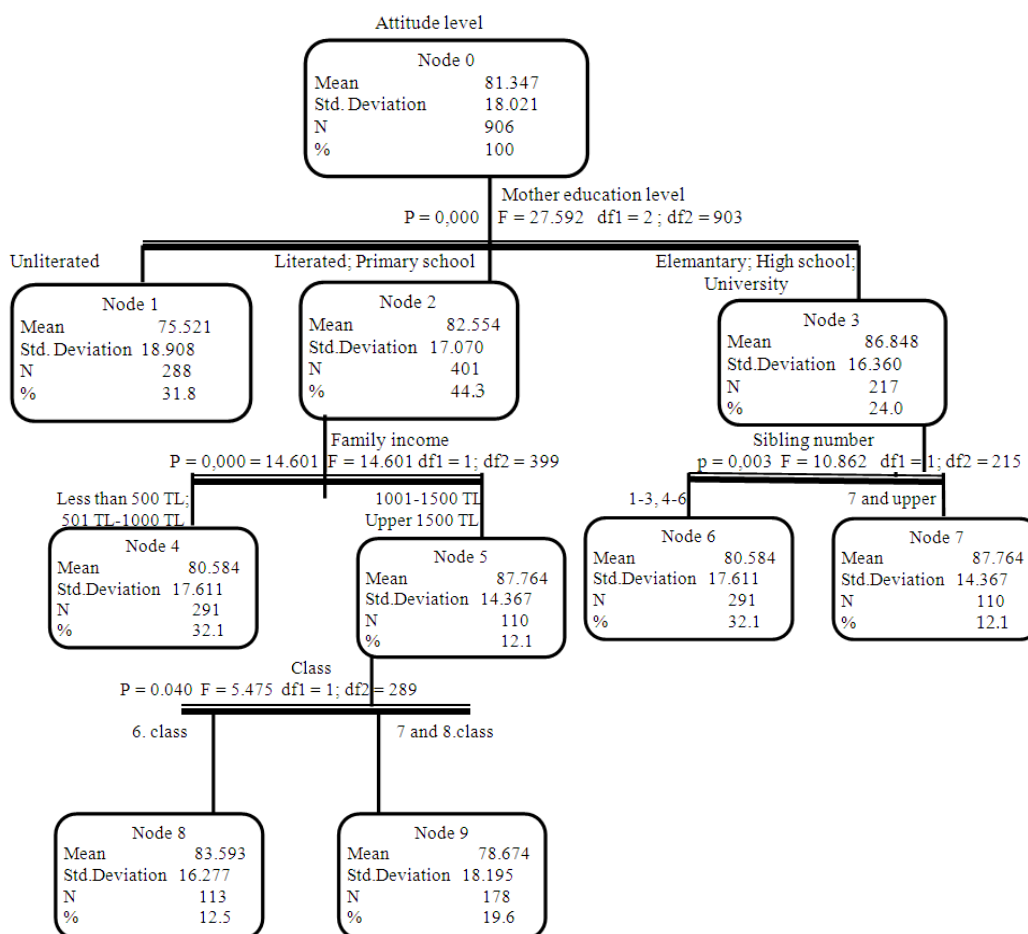
Fig. 3: Result of Chaid analyze related to first data set (without ANN)

level had a significant effect on the students' attitudes whose mothers were "literate and primary school graduates". The number of siblings had a significant effect on the students' attitudes whose mothers were secondary school graduates, high school graduates or university graduates. The tree shaped diagram of "income level", which was considered significant in the model, showed that the variable interacted with "grade".

As it is clear from Fig. 4, unlike the previous model, "educational background of father" had an overwhelming effect on the students' attitudes in the data set estimated by ANN. Another variable in interaction which was considered significant in the model was "educational background of mother". However, it was observed that educational background of mother had a significant effect on the students' attitudes whose fathers were "illiterate", "literate" or "primary school graduates". The number of siblings had a significant effect on the students' attitudes whose

fathers were secondary school graduates, high school graduates or university graduates. Again, educational background of mother had a significant effect on the students' attitudes in the number of siblings range from 1-3.

DISCUSSION

Main aim of the research was to show use of Artificial Neural Networks for parameter estimation. The initial data set obtained by the measurement tool and the new data set obtained by Multi Layer Perceptron Algorithm, which had the highest estimation probability for the research, were individually examined by Chaid analysis and the results were compared. Analysis for the initial data set showed educational background of mother had a significant effect on the dependent variable of total score. In the research, it was observed that the income level had a significant effect on the students' attitudes whose
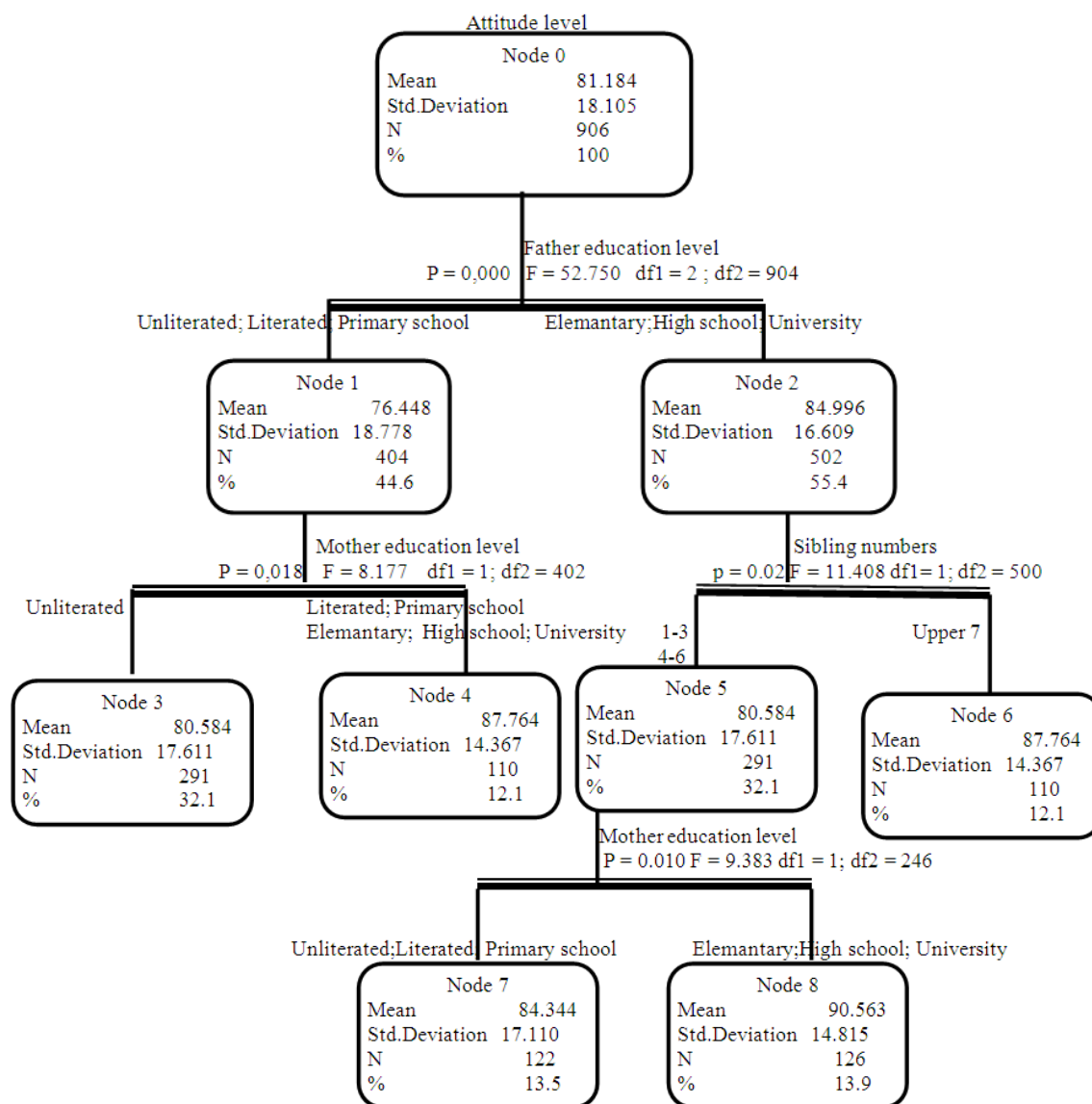
Fig. 4: Result of chaid analyse with ANN

mothers were "literate" or "primary school graduates", while the number of siblings had a significant effect on the students' attitudes whose mothers were secondary school graduates, high school graduates or university graduates. The tree diagram of "income level", which was considered significant in the model, showed the students interacted with grade. Unlike the previous model, analysis for the data set estimated by ANN showed educational background of father had a significant effect on the students' attitudes. Another variable in interaction which was considered significant in the model was "educational background of mother". Yet, educational background of mother had a

significant effect on the students' attitudes whose fathers were "illiterate", "literate" or "primary school graduates". The number of siblings had a significant effect on the students' attitudes whose fathers were secondary school graduates, high school graduates or university graduates. Again, educational background of mother had a significant effect on the students' attitudes in the number of siblings range from 1-3.

CONCLUSION

As a result, most studies in the literature suggest models by ANN are more consistent (Badr *et al.*, 2003;

Durmuş and Meric, 2005; Kurt and Ture, 2005; Uzun and Erdem, 2005; Wieland and Mirschel, 2008). The findings of the research show Artificial Neural Networks could be used for parameter estimation in cause-effect based studies in educational sciences. It is also thought the research will contribute to extensive use of advanced statistical methods.

REFERENCES

Ayhan, T., 2009. Artificial Neural Network Lecture Notes. 2009,http://www2.itu.edu.tr/~ayhant/dersler/ysa/ysa.html

Badr, E.A., G.E. Nasr and C. John, 2003. Backpropagation neural networks for modeling gasoline consumption. Energy Converse. Manage., 44: 893-905.

Durmuş, K.H. and C. Meric, 2005. Using artificial neural network (ann) in mechanical engineering. Eng. Mech. (Muhendis ve Makine), 46: 47-56.

Erilli, N.A., E. Egrioglu, U. Yolcu, C.H. Aladag and V.R. Uslu, 2010. Inflation with hybrid of feed forward and recurrent artificial networks. Doguş Univ. J., 11: 42-55.

Elmas, C., 2003. Artificial Neural Network. Ankara, Seckin Publishing, ISBN: 975 347 612 4, pp: 21.

Fausett, L., 1994. Fundamentals of Neural Networks. Prentice Hall, ISBN: 0133341860, pp: 461.

Gardner, M.W. and S.R. Dorling, 1998. Artificial neural networks (the multilayer perceptron)-A review of applications in the atmospheric sciences. Atmospheric Environ., 32: 2627-2636.

Golden, R.M.,1996. Mathematical Methods for Neural Network Analysis and Design. 1st Edn., Massachusetts Institute of Technology, USA., ISBN: 0-262-07174-6, pp: 411

Hamzacebi, C. and F. Kutay, 2004. Electric consumption forecasting of Turkey using artificial neural networks up to year 2010. J. Fac. Eng. Arch. Gazi Univ., 19: 227-233.

Haykin, S., 1999. Neural Network: A Comprehensive Foundation. 2nd Edn., Upper Saddle River NJ. Prentice-Hall, New York, ISBN: 0780334949, pp: 51.

Karasar, N., 1991. Scientific Search Method. Ankara, Nobel Yayınları, ISBN: 9755910468, pp: 156.

Kayri, M. and M. Boysan, 2007. Using chaid analysis in researches and an application pertaining to coping strategies. Ankara Univ. J. Faculty Educ. Sci., 40: 133-149.

Kurt, I. and M. Ture, 2005. Comparison of artificial neural networks and logistic regression analysis in determining factors affecting alcohol consumption among medicine students. Trakya Univ. Tıp Fak Derg., 22: 142-153.

Manel, S., J.M. Dias and S.J. Ormerod, 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: A Case study with a Himalayan river bird. Ecolo. Model., 120: 337-347.

Schalkof, R., 1992. Pattern Recognation: Statistical. Structural and Neural Approaches. Wiley, ISBN: 0471529745, pp: 154.

Uzun, E. and O.A. Erdem, 2005. Turkish times new roman, arial, and handwriting characters recognition by neural network. J. Fac. Eng. Arch. Gazi Univ., 20: 13-19.

Wieland, R. and W. Mirschel, 2008. Adaptive fuzzy modeling versus artificial neural networks. Environ. Model. Software, 23: 215-224.

Yılmaz, I.H., 2009. The students' current expectations from their families in their learning process and their opinions about the expected outcome. Unpublished Master Thesis, Yuzuncu Yil University, Institue of Social Science, Van/Turkey.