Prediction of Success or Fail of Students on Different Educational Majors at the End of the High School with Artificial Neural Networks Methods

Sayyed Majid Mazinani, *Member, IACSIT*, and Sayyede Azam Abolghasempur

***Abstract*—The main objective of departments of education is providing quality education for students. One of the methods to achieve a high level of quality in education system is knowledge discovery for prediction of students' enrollment in a specific field. This knowledge is hidden in educational and motivational data. It is extractable by data mining techniques. The present article is designed to justify the capabilities of data mining methods in education by presenting a data mining method. In this research classification is used for evaluating the students' performance. There are many methods for predicting students' Performance in different fields. Here artificial neural (ANN) is used. With this method we extract the knowledge which describes the students' performance in the final exams. This method is suitable for identifying appropriate field for the students and they will succeed in the field.**

***Index Terms*—Activation function, artificial neural networks, data mining training, multilayer perceptron networks.**

I. Introduction

Data mining is the process which was appeared in 1990's and it was introduced by Fayaz for the first time in the first international conference on knowledge discovery and data mining in 1995. In 1959 the term Machine Learning was introduced for the first time by Samuel and it refers to the ways in which a computer can gain knowledge directly from data. Pendy and Pal conducted their study to students' performance by Selecting 600 students from different faculties [1]. Then Hejazi and Naghavi have done research on the student's attitude toward class attendance, study hours after university and etc. They used simple linear regression analysis. In this research we intended to use a method in which students can choose a field easier and with less anxiety and the probability of success will be greater at that field. This is the most important matter for the students and parents. The technique used in this research is artificial neural network in data mining and Matlab software was used for simulations [2].

II. Artificial Neural Network Structure

An artificial neural network consists of elements of layers and weights. The network behavior is related to the connections between members. Generally, there are three types of neuron layers in neural networks. Input layer, hidden layers, output layer. There are single and multilayer networks. Single layer organization in which all units are connected to one layer is the most common type. In multilayer networks, units are numbered according to layers (instead of tracking overall numbering).Two layers of a network are connected with weights and in fact with connections [3]. There are several types of connections and weight connections in neural networks.

- *Forward:* Signals move only in one direction. Each layer output has no effect on that layer.
- *Backward:* Data are refereed from upper layer nodes to lower layer nodes.
- *Lateral:* Each layer output is used as the nodes input of the same layer. *Final Stage*

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables.

A. Multilayer Perceptron Networks

In multilayer perceptron networks at first weights of the previous layers will be corrected [3].

Perceptron networks consist of an input layer, several hidden layers, and an output layer. The layers have the following conditions:

- Each layer's neurons are only connected to the next layer neurons.
- Each Neuron is connected to all neurons of the next layer.
- Input layer neurons do no action and their weights are constant and are equal to one.
- These neurons have no compression function. The operator propagation is forward. All neurons except input layer are collecting neurons and each neuron can have an independent compression function.
- Each neuron can have independent bias.
- The number of hidden layers is not known.

B. Learning Algorithm of Multilayer Perceptron Networks (Error Propagation)

Signs and names

In this learning algorithm the following signs are used [4]: x training input vector t target output vector

δ_k the section for justifying weights for w_{jk} and the value of that is calculated according to error input unit(the difference between the network output and the target output)

Manuscript received June 10, 2013; revised August 20, 2013.

Sayyed Majid Mazinani is with the Department of Electrical and Computer Engineering, Imam Reza University, Mashhad, Iran (email: mazinani@ieee.org).

Sayyede Azam Abolghasempur is with the Science and Research Unit, Islamic Azad University, khorasan Razavi, Iran (email: a.abolghasempur@yahoo.com).

DOI: 10.7763/IJIMT.2013.V4.442

δ_i the section for justifying the error correction weights for v_{ij} which is calculated after error propagation from output layer to the hidden unit z_j α learning rate

x_i input unit of i v_{0j} bios in hidden unit j

z_j hidden unit j

Input unit z_j which is shown by z_in_j

z_in_j is shown in (1)

$$z\_in_j = v_{0j} + \sum_i X_i v_{ij} \tag{1}$$

Output signal activation of z_j is named z_j and it is calculated according to (2) in which f is activation.

$$z_j = f(z_{in_j}) \tag{2}$$

C. Activation Function

Activation function in neural networks must be continuous, differentiable, and the derivative must be univocal descending.

One of the most common functions is the sigmoid dual function.

Domain of this function is from 0 to 1 and the formula is like (3)

$$f = \frac{1}{1+\exp(-x)} \tag{3}$$

Another activation function is bipolar sigmoid function. This function domain is -1 to 1 and the formula is as the (4).

$$f = \frac{2}{1+\exp(-x)} - 1 \tag{4}$$

D. Learning Stages of Error-Back Propagation Algorithm

- Stage 0: We give an initial value to weights and biases (very small accidental value)
- Stage 1: We continue stages 2 and 9 until the stopping conditions are established
- Stage 2: For each training pair (target and input values) we do stages 3 and 8

1) Feed forward

- Stage 3: Each input unit x_i, i=1, …, n receives input signal x_i and spreads it to all units in the next layer.
- Stage 4: Each hidden layer z_j, j=1, …, p collect it's weighted input signal, according to (5)

$$z\_in_j = v_{0j} + \sum_{i=1}^{n} x_i v_{ij} \tag{5}$$

Then the activation function is used for calculating the output, according to (6) formula and send the signal to all next units.

$$z_j = f(z_{in_j}) \tag{6}$$

- Stage 5: Each output unit Y_k, k=1, …, m collects It's weighted input signal according to (7)

$$y\_in_k = w_{0k} + \sum_{j=1}^{p} z_j w_{jk} \tag{7}$$

And then the activation function is used for calculating the output, according to (8)

$$y_k = f(y_{in_k}) \tag{8}$$

2) Error propagation

- Stage 6: Each output unit Y_k, k=1, …, m receives a target model which is corresponding to input training model and calculate the error according to (9) formula

$$\delta_k = (t_k - y_k) f'(y\_in_k) \tag{9}$$

And then the weight correction parameter is calculated which will be used in w_{jk} update according to (10) formula

$$\Delta w_{jk} = a\delta_k z_j \tag{10}$$

And then the biases correction parameter is calculated which will be used in w_{0k} update according to (11) formula

$$\Delta w_{ok} = a\delta_k \tag{11}$$

And then send δ_k to previous unit later.

- Stage 7: Each hidden layer collect it's delta outputs, according to (2-12) formula

$$\delta\_in_j = \sum_{k=1}^{m} w_{jk}\delta_k \tag{12}$$

Then it is multiplied to activation function derivative to calculate error data parameters. It is calculated according to (13) formula:

$$\delta_j = \delta\_in_j f'(z\_in_j) \tag{13}$$

Then weight correction is calculated which in next stages will be used for updating v_{ij} according to (14) formula:

$$\Delta v_{ij} = a\delta_j x_i \tag{14}$$

And then bios correction is calculated which later will be used for updating v_{0j} according to (15) formula:

$$\Delta v_{0j} = a\delta_j \tag{15}$$

3) Weight and bios update

- Stage 8: Each output unit updates It's weights and biases according to (16) formula:

$$j = 0,\ldots,p$$
$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk} \tag{16}$$

Each hidden unit z_j, j=1, …, p updates its weight and biases according to (17) formula:

$$i = 0,\ldots,n$$
$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij} \tag{17}$$

- Stage 9: We check stopping conditions. We should be careful that in implementing this algorithm, separate arrays must be used for output units Deltas (Delta in stage 6) and hidden units Deltas.

E. Final Condition and Early Weights Helpful Hints

A period is a complete cycle around learning complex. Several cycles are needed for training a back propagation neural network. Usually, training cycle is continued to get a total error average with suitable least amount or zero (final condition). Sometimes, average error doesn't change in a few periods. So that neural network learning is finished.

Generally, early extent of weighs and biases are done by little extent and randomly. It has been confirmed that choosing early extent for weighs and biases converge to correct amounts. That is, if each extent is chosen, network parameters are regulated. The difference is that, while much distance to early weighs and biases is observed with their correct mounts, period numbers increase [5], [6].

When there is more than one layer in perceptron networks, the algorithm should be generalized to all layers. There is not an available practical method for estimating unit numbers (neurons) in each hidden layer. Therefore, trial and error methods should be used to get a suitable amount of error average.

F. Radial Basis Function Networks

Through using (18) equations, radial basis function network can be calculated and only activating function should be focused and the parts of the algorithm are the same as previous part. That is, algorithm is used for measuring weighs and biases after error propagation [7]. In above equation, || || is the Euclidean distance.

$$\begin{aligned}
& j = 0,1,\ldots,k; i = 0,\ldots,n; l = 1,\ldots,m \\
& Znet_j = \sum_{i=0}^{k} w_{ij} x \\
& Z_j = f_1(Znet_j) \\
& Ynet_l = \sum_{j=0}^{k} w_{ij} z_j \\
& Y_l = f_2(Ynet_l) \\
& Y_l = \sum_{j}^{k} f_j w_{jl} \\
& f_j = f(\| x - z_j \|)
\end{aligned} \tag{18}$$

III. Applied Model in the Study

In this study multi-layer neural networks have been used for predicting scientific and motivational progresses among studied in various fields of study. As a matter of fact, the used model for this predicting includes an input, a hidden layer and an output. After conducting the inputs and outputs to neural network (neural network training), the neural network learns to create a suitable corresponding reply to the same inputs by taking the inputs again.

A. Inputs

The inputs contain two parts, course grades and educational talent tests; both of them are expressed numerically. In course grades, the scores of all textbooks in first grade of high school are gathered for all students and in educational talent, the psychological tests, asked from students, are given.

The inputs of textbooks scores are as follows (all between 0 to 20):

Religious and life, Persian language, Persian literature, Arabic language, English language, physics and laboratory, chemistry, math, social studies, science, sport, life skills, educational and occupational planning.

The inputs of educational talent include the following questions:

- The extent of liking or hating the field of study by the student (-1 to 1), while -1 indicates that the student hates the field and 1 means the students likes the field very much.
- The extent of encouraging or forcing to choose a field for the student by parents (a number between -1 to 1), while -1 indicates the student has to choose the field just because of parents forcing and number 1 means that the study chooses the field as his/her interests.
- The extent of facilities and schools (-1 to1), while number -1 means that students in the field of study have the east educational facilities and number 1 means that this field contains the most educational facilities.

B. Outputs

Our output in this study is just a number that is there is just a neuron in out layer that corresponds to total average of students. In fact, this number is obtained through conducting a non- linear function (neural network) on the inputs.

Total average is output, in fact output, students' total average, is predicted by using neural network and its corresponding inputs.

IV. Simulations and Evaluation

The supposed neural network looks like as following figure in which the inputs are the scores and educational talent tests and the outs pub is the total average.

Back propagation of error has been used for training the network and also Euclidean activating function is used for creating a Radial Basis Network model. General scheme of a multilayer neural network is given in Fig. 1 [8].

A. Conducting the Data and Errors in Proposed Neural Networks

60% information of data bases are used for training neural networks and 40% for validating and testing the estimated figures.

The process includes entering the data, dividing them into two training and testing parts, preprocessed normalizing, regulating hidden layers training function and other parameters, training neural network, confirming and testing neural network.

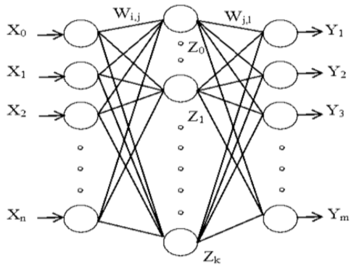

Fig. 1. Multi layers neural network.

B. Normalizing Inputs and Outputs

Diverging and incorrectly training the network decreases and training neural networks will be quicker (less time or less repetitions) by scaling inputs and outputs , in scaling the inputs and outputs, the data are normalized in [-1, 1] (bipolar). Due to (19):

$$x_{norm} = 2\left(\frac{x_{real} - x_{\min}}{x_{\max} - x_{\min}}\right) - 1 \tag{19}$$

where, x_{norm} is normalized extent of the parameter, x_{real} is the real parameter extent, $x_{\min}$ is the least parameter extent and $x_{\max}$ is the highest parameter extent. Also the data are divided randomly to increase validity in this process.

C. Error Measurement

Mean squared error (MSE) is used as a tool in measuring predicting precision. Due to (20):

$$MSE = \sum_{i=1}^{N} \frac{(P_{real} - P_{\min})^2}{N} \tag{20}$$

where, N is sample numbers, P_{actual} is real or actual extent of parameter and P_{sim} is simulated extent of the parameter [8].

V. Simulations

The data in various fields of study (math, humanities, experimental, technological and occupational) are given to neural network for training and examination and in fact, we have 5 neural networks which training them through specialized filed data. The data were gathered actually and from schools in various areas. The used topology in this study is given in Fig. 2.

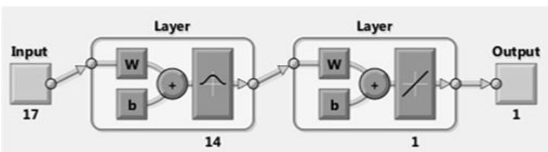

Fig. 2. Multi layers neural network.

Fig. 3, Fig. 5, Fig. 7, Fig. 9, and Fig. 11 show averages of student and Fig. 4, Fig. 6, Fig. 8, Fig. 10, and Fig. 12 show error square average. Here, the results of simulation for one field are given:

Output error is 10-31 for math field that is suitable and are given in Fig. 3 and Fig. 4.

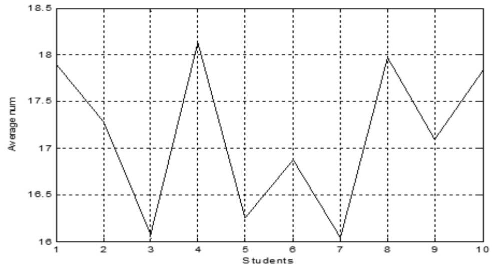

Fig. 3. Output averages for many students.

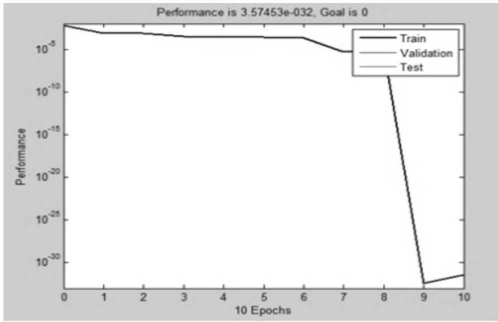

Fig. 4. error square average.

Output error is 10-32 for experimental field that is suitable and are given in Fig. 5 and Fig. 6.

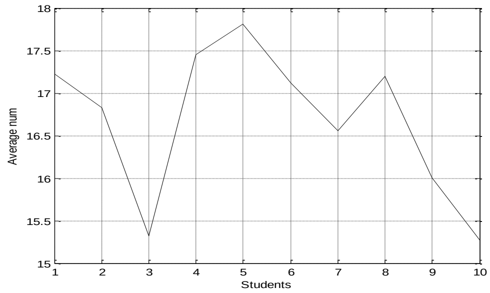

Fig. 5. Output averages for many students.

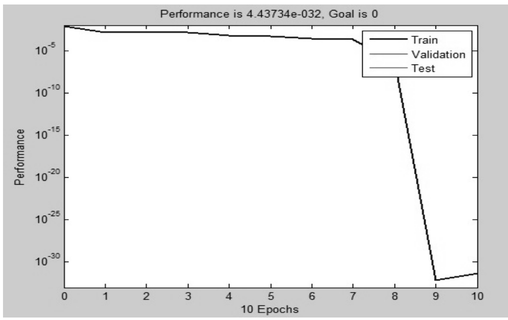

Fig. 6. error square average.

Output error is 10-32 for humanities field that is suitable and are given in Fig. 7 and Fig. 8.

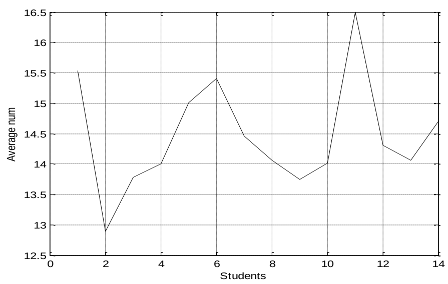

Fig. 7. Output averages for many students.

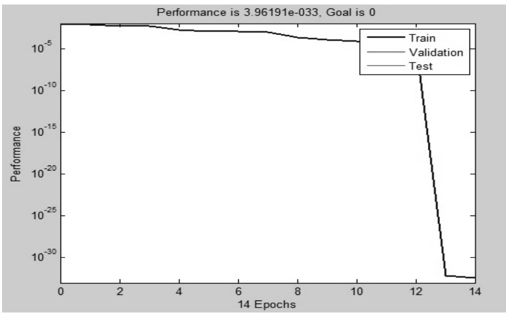

Fig. 8. Error square average.

Output error is 10-31 for technological field that is suitable and are given in Fig. 9 and Fig. 10.

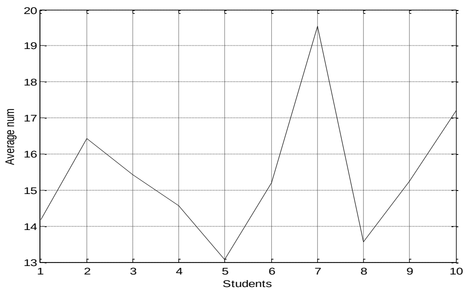

Fig. 9. Output averages for many students.

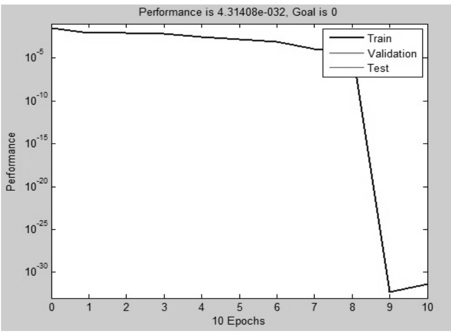

Fig. 10. Error square average.

Output error is 10-32 for occupational field that is suitable and are given in Fig. 11 and Fig. 12.

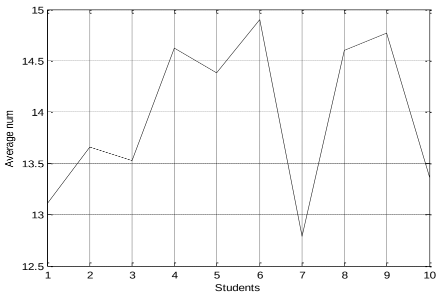

Fig. 11. Output averages for many students.

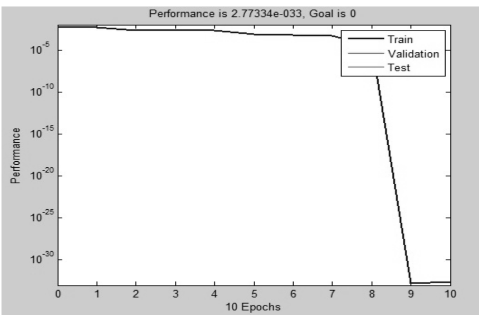

Fig. 12. Error square average.

Systems neural networks and novel measuring methods for mechanical learning, indicating knowledge and finally conducting the obtained knowledge are going to conduct output replies of complicated systems. In fact, neural networks are non- linear tools in predicting various data.

VI. Conclusion

In this study, we predict scientific and motivational progress among students in various field of study through using artificial neural networks; regarding to this method the scores of first grade high school and some educational talent and motivational questions was considered as inputs. Error back propagation method with Euclidean activating function has been applied for training neural network leading to Radial Basis Networks. Finally, the data have been examined on 5 neural networks, everyone has been a fired of study. The networks could get the error less than 10-31 that is little and suitable. And eventually the 5 trained networks were examined with one ideal input and average results are given to them; that proposes the students to choose which filed for continuing studies; in other words in which they will be more successful.

References

[1] DANN: Genetic wavelets, *DANN Project*, August 2010.

[2] Springer, doi: 10.1007/978-3-642-01513-7_6, ISBN 978-3-642-01215-0, January 2012.

[3] NASA Neural Network Project Passes Milestone, NASA, July 2010.

[4] D. C. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-Column deep neural network for traffic sign classification," *Neural Networks*, 2012.

[5] I. E. Livieris and P. Pintelas, "An improved spectral conjugate gradient neural network training algorithm," *International Journal on Artificial Intelligence Tools*, vol. 21, no. 1, 2012.

[6] C. Romero, S. Ventura, and E. Garcia, "Data mining in course management systems: Moodle case study and tutorial," *Computers and Education*, vol. 51, no. 1, pp. 368–384, 2008.

[7] I. E. Livieris and P. Pintelas, "An improved spectral conjugate gradient neural network training algorithm," *International Journal on Artificial Intelligence Tools*, vol. 21, no. 1, 2012.

[8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

Sayyed Majid Mazinani was born in Mashhad, Iran on 28 January 1971. He received his Bachelor degree in Electronics from Ferdowsi University , Mashhad, Iran in 1994 and his Master degree in Remote Sensing and Image Processing from Tarbiat Modarres University, Tehran, Iran in 1997. He worked in IRIB from 1999 to 2004. He also received his phD in Wireless Sensor Networks from Ferdowsi University , Mashhad, Iran in 2009. He is currently assistant professor at the faculty of Engineering in Imam Reza University, Mashhad, Iran. He was the head of Department of Electrical and Computer Engineering from 2009 to 2012. His research interests include Computer Networks, Wireless Sensor Networks and Smart Grids.

Seyede Azam Abolghasempur is from Iran and lived in Bojnourd. She was born in 1984, and got her B.Sc on software engineering in 2009. Then she started teaching at Training and Educational organization in high schools as well as in university. She got first location on electronic-content- producing between all Training and Educational offices in Iran in 2011. She was accepted on Research and Science of Khorasan-Razavi in M.sc k in 2011, which was support by Training and Educational organization.