ANALYZING THE FACTORS AFFECTING THE SUCCESS IN UNIVERSITY ENTRANCE EXAMINATION THROUGH THE USE OF ARTIFICIAL NEURAL NETWORKS

Zafer Ağdelen, Ali Haydar, Andisheh Kanani
Girne Amerikan University, Turkish Republic of Northern Cyprus
zagdelen@gau.edu.tr, ahaydar@gau.edu.tr

ABSTRACT

There are many factors that affect the success of students in university entrance examination. These factors can be mainly categorized as follows; social factors, environmental factors, economical factors etc. The main aim of this study is to find whether there is a relation between these factors and the success in the university entrance examination. To achieve this goal, we use radial basis functions (RBF) network which is one of the artificial neural networks approach used to find a relation between input and output variables nonlinearly. From the experimental results, we observed that the relation between the input variables that corresponds to social, economical and environmental factors and the output variable can be modeled nonlinearly using radial basis functions network.

Keywords: Artificial neural networks, success, education, entrance examination

INTRODUCTION

In developing countries, university education has a great importance in finding better job opportunites. In Turkey, in order to enter a university, there is a single central examination to allocate the students to the related disciplines. Since the capacity of the universities providing higher education is limited, only approximately one tenth of the students who enter this examination succeed. Because of this reason, there is a big competetion among students and it is very important to get a good mark from the entrance examination. In order to be successful in this examination,there are numerous factors that affect the result taken from the examination. It is obvious that the main component in this success is intelligence and regular study of a student for this examination. Apart from thesefactors, there are several other factors that may affect the performance of a student. These factors can be categorized as social, environmnetal, economical etc. In literature, there are several studies aiming to analyze the relation between academic success and the factors mentioned above (Büyüköztürk and Deryakulu, 2002; Begik, 1997; Özgüven, 1974; Parke and Welsh; 1998; Perçin , 1998).

Most of the studies about the impact of the factors on academic success of students attempted to estimate the statistical relationship linearly. However, the relationship between factors and academic success are not examined by using nonlinear techniques.

One method for this nonlinear approach that can be used is Artificial Neural Networks (ANN's). Artificial Neural Networks are systems that are deliberately constructed to make use of some organizational principles resembling those of the human brain (Lin and Lee, 1996). We decided to use one of the modeling technique of ANN approach which is called as Radial Basis Function in this work. In this study, our aim is to check whether the relation between several factors and university examination success can be modelled nonlineraly or not by means of the Artificial Neural Network modeling technique.

METHOD

Participants

The population of the research consists of all freshman students studying in Girne American Univertsity. A sample of 206 were collected from the students and only 154 of them were considered to be used in analysis. 52 of the questionnaires were not answered properly, hence they were excluded from the study. From these collected data some of them were selected randomly to be used as training data and the rest were used for testing our model.

Apparatus

In order to analyze the relation, a questionnaire was structured to collect data. This questionnaire used in this study consists of three sections. In the first section several questions related with the social factors that can affect the education were asked. In the second section some questions were asked which aims to see the effect of economical conditions on success. In the third section, questions related with the environmental factors were prepared. Also the point taken in university entrance examination were asked.

MODEL DESCRITION

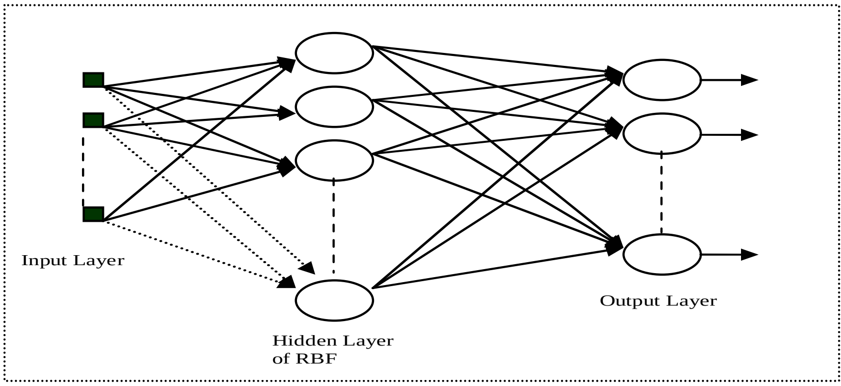

Figure 1: Radial Basis Function Network

Radial Basis Function Networks are a type of artificial neural networks that can be viewed as a curve-fitting problem in high dimensional space. According to this viewpoint, learning is equivalent to finding a surface in hyperspace that provides the best fits to the training data. The ‘best fit’ is measured in some statistical sense. The RBF network is able to approximate a wide class of nonlinear multidimensional functions (Chen, Corun and Grant, 1991; Chen, Gibson, Corun and Grant, 1991).

The construction of a radial-basis function network, in its most basic form, involves three layers with entirely different roles (Haykin, 1999; Cichoclar and Unbehaven, 1993). The input layer is made up of nodes that connect the network to its enviroment. The second layer is the only hidden layer in the network. The third layer, which is the output layer, supplies the response of the network for the applied signal at the input layer. The RBF network used in this work is given in Figure 1. It consists of an input layer, one hidden layer and an output layer.

The transformation from input space to output is nonlinear and the transformation from hidden unit space to output space is linear. the set of Basis functions$\{\Phi_i(x)|i = 1,2,........,M\}$ is defined as follows

$$\Phi_i(x) = G(||x-t_i||) = \exp(-||x-t_i||),$$

$$i=1,2,3,.......,M \qquad (1)$$

Where $\{t_i|i=1,2,.......,M\}$ are the set of M centers to be determined and x is one of the training(input) data in a set $\{x_i|i = 1,2,3,.........,N\}$ of size N. Typically ,the number of basis functions, M is less than the number of data points (i.e $M \leqslant N$). Our aim is to find the suitable w values in order to minimize the Euclidean Norm

$$|| d-Gw ||^2 \qquad (2)$$

where

$$d=[d_1,d_2,..........,d_N]$$

$$G=\begin{bmatrix} G(||x_1-t_1||) & G(||x_1-t_2||) & \ldots & G(||x_1-t_M||) \\ G(||x_2-t_1||) & G(||x_2-t_2||) & \ldots & G(||x_2-t_M||) \\ . & . & & . \\ . & . & & . \end{bmatrix}$$

$G(\|x_N - t_1\|) \quad G(\|x_N - t_2\|) \quad \ldots.. \quad G(\|x_N - t_M\|)$

$w = [w_1, w_2, \ldots\ldots\ldots.., w_M]^T.$

The vector d is an N-dimensional desired response vector, the matrix G is an N×M matrix of Green's function and the vector w is M-by-1 weight vector for the linear transformation from hidden unit space to output space. The minimum norm solution to the over determined least squares data fitting problem can be given as follows

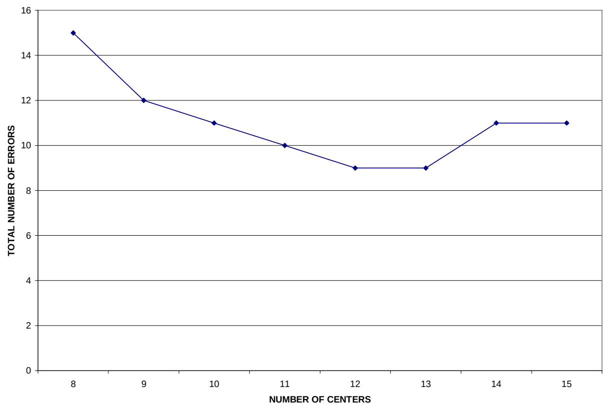

Figure 2.:Total Number Errors According to the Number of Centers Used in RBF

$$w = (G^T G)^{-1} G^T d \qquad (3)$$

The set of centers $\{t_i | i=1, 2, 3\ldots, M\}$ can be selected randomly from the set of data points, can be selected using the clustering techniques to find the suitable centers or can be selected using gradient descent algorithm. In this study, we used random selection to find the set of centers for the Radial Basis Functions.

RESULTS

In this paper, the main aim is to observe whether there is a relation between the social, environmental, economical factors and the success in the university entrance examination or not using one of the artificial neural networks approach which is called as radial basis functions network.

The data, that are used in this work, are collected from 154 students, who are studying in the first year in GAU, by means of a questionnaire. These students are selected from different faculties but mostly they are from the engineering and architecture faculty.

In this study, a total of 21 performance measures obtained from the questionnaire are used. From these 21 variables, 20 of them are used as an input variables and one of them are used as an output variable. The output variable is representing the point taken from the University entrance examination and the 20 input variables are mainly the multiple choice questions that are related with the factors mentioned above. The data is divided into 5 different groups by selecting 5 intervals of the point taken from the university entrance examination in such a way that each group will approximately contain the same amount of data. Hence there are 20 input variables to the RBF network and 5 outputs that shows the intervals of the output variable.

Before using the input variables in RBF network, all of them are normalized. The normalization is mainly performed to equate the effect of variables in our model.

First of all, an experiment is performed in order to select a suitable number of centers, M, for the radial basis function network. In order to achieve this goal, all the selected 154 student data for this study are used as a training data. The number of wrongly classified training data for different number of centers are shown in Figure 2.

From this figure, even though the total training errors are very close, we decided to fix the number of centers M to 12 and from this point on, all the results are obtained using this fix number 12. The next step is to see whether the model at hand is capable of establishing a relation between the input variables and an output variable in such a way that it will correctly classify the output variable. To achieve this goal, the data is divided into two as training data and testing data. After the training of the model using the training data, test data is applied to see the performance of our model. Table 1 shows the percentage correct classification of the test data when the data is divided randomly into training and testing part as shown below.

A similar experiment is performed one more time as it was mentioned above and tabulated in Table 2. The only difference between the two set of results shown in Table 1 and Table 2 is that the number of training data for each class is randomly decided in Table 1 however the number of training data for each class is equated for the second experiment.

Table 1: Percentage Error Obtained As the Numnber of Test Data Changes (Training Data Selected Randomly)

Number of Train Data	Number of Test Data	Number of Error	Percentage Error (%)
120	34	5	14.7
115	39	5	12.8
110	44	6	13.6
100	54	7	13.0
90	64	14	21.9
85	69	18	26.1
80	74	22	29.7
70	84	31	37

Table 2: Percentage Error Obtained As the Number of Test Data Changes

Number of Train Data	Number of Test Data	Number of Error	Percentage Error (%)
120	34	3	8.8
115	39	4	10.3
110	44	4	9.1
100	54	6	11.1
90	64	10	15.6
85	69	15	21.7
80	74	18	24.3
70	84	24	28.6

CONCLUSION

The University entrance examination that is held once a year in Turkey is very important for the students in the sense that it shapes the future life of the students. In order to be successful in this examination, there are many factors that affect the result taken from the examination.

In this paper, our aim is to find whether the relation between some of these factors and the point taken from the University entrance examination can be modeled using a nonlinear model or not. To check this claim we use one of the modeling technique, namely the Radial Basis Functions. This technique is one of the approaches used in Artificial Neural Networks for representing any kind of relation. From our experimental results, our conclusion can be summarized as follows;

- By using the Radial Basis Functions Netwok in the modeling of this relation, we observed that nonlinear model is capable of representing the relation at hand.
- From the results obtained in Table 1 and in Table 2, it is clear that the correct classification in percentages increases as the number of traing data increases and it becomes steady when an enough number of training data is used.
- Comparing the results obtained in Table 1 and Table 2, one can easily observe that the selection of equal number of training data from each class increases the correct classification.
- This model can be used as an estimator of the output variable if the input variables

that corresponds to the factors mentioned above are known. From this observation, one can conclude that these factors affect the performance of a student in the university entrance examination and through the improvement of these factors in positive manner, one can achieve better results in this examination.

REFERENCES

Begik, Ş. (1997). **Akademik Başarının Kestirilmesinde Yetenek, İlgi, Kişilik, Lise Başarısı ve ÖSYS Başarısının Etkisi**. Marmara Üniversitesi Eğitim Bilimleri Enstitüsü (Yayımlanmamış Yüksek Lisans Tezi).

Büyüköztürk, Ş. And Deryakulu, D. (2002). **Bilgisayar ve Öğretim Teknolojileri Öğtremenliği İle Sınıf Öğretmenliği Programı Öğrencilerinin Akademik Başarılarını Etkileyen Faktörler**. Kurum ve Uygulamada Eğitim Yönetimi Dergisi. Sayı 30. sayfa:187-204.

Chen, S., Corun, S.F.N. and Grant, P.M. (1991), **Orthogonal least square learning algorithm for radial basis function networks**, IEEE Trans, Neural Networks, Vol. 2, pp. 302-309.

Chen, S., Gibson, G.J., Corun, C.F.N, and Grant, P.M. (1991), **Reconstruction of binary signals using an adaptive radial-basis-function equilizer,** Signal Processing, Vol. 22, pp. 72-93

Cichoclar, A. and Unbehaven, R. (1993), **Neural Networks for Optimization and Signal Processing**, John Wiley & Sons.

Haykin, S. (1999), **Neural Networks**, second edition, Prentice Hall.

Lin, C. T. And Lee, C.S. (1996), **Neural Fuzzy Systems,** Prentice Hall.

Parke, R. and Welsh, M. (1998). **Social Relationship & Academic Success**. Trust For Educational Leadership. 28 (1). pp: 32-34

Perçin, A. (1998). **Karadeniz Teknik Ünversitesi (KTÜ) Lisans Öğrencilerinin Başarı Düzeyini Etkileyen Faktörlerin Kantitatif Analizi**. Karadeniz Teknik Üniversitesi Sosyal Bilimler Enstitüsü (Yayımlanmamış Doktora Tezi).

Özgüven, İ.E. (1974). **Üniversite Öğrencilerinin Akademik Başarılarını Etkileyen Zihinsel Olmayan Faktörler**. Hacettepe Üniversitesi Yayınları. Ankara