World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:4, No:10, 2010

Forecasting e-Learning Efficiency by Using Artificial Neural Networks and a Balanced Score Card

Petar Halachev

***Abstract*—**Forecasting the values of the indicators, which characterize the effectiveness of performance of organizations is of great importance for their successful development. Such forecasting is necessary in order to assess the current state and to foresee future developments, so that measures to improve the organization's activity could be undertaken in time. The article presents an overview of the applied mathematical and statistical methods for developing forecasts. Special attention is paid to artificial neural networks as a forecasting tool. Their strengths and weaknesses are analyzed and a synopsis is made of the application of artificial neural networks in the field of forecasting of the values of different education efficiency indicators. A method of evaluation of the activity of universities using the Balanced Scorecard is proposed and Key Performance Indicators for assessment of e-learning are selected. Resulting indicators for the evaluation of efficiency of the activity are proposed. An artificial neural network is constructed and applied in the forecasting of the values of indicators for e-learning efficiency on the basis of the KPI values.

***Keywords*—**artificial neural network, balanced scorecard, e-learning

I. Forecasting as a Means of Increasing Efficiency

THE successful realization of high schools on the educational services market depends on their efforts to attain the defined goals. Universities are not closed systems and their activity is influenced by many external factors, whose effect is difficult to anticipate. Different forecasting methods are used for this purpose. The most popular among them are the following:

•extrapolation, based on the hypothesis that the existing relationships will continue to influence events during the forecast period. In this case, the forecasted values are calculated as a dynamic row extension in the future, according to a manifested relationship. Extrapolation can be presented as a correlation and regression relationship or in the form of a trend, defined on the basis of factor analysis.

•for long-term forecasts, in cases when it is difficult to make a quantitative evaluation of indicators, it is possible to use the expert assessment method. Specialists in some thematic field develop long-term forecasts using individual or group expert assessment, the interview method, the expert committee method, etc.

P. H. Author is with the Institute of Information and Communication Technologies, Bulgarian Academy of Science, Sofia, acad. G. Bonchev str. Block 2, 1113, Bulgaria (e-mail: halachev@iit.bas.bg).

•the analogy method transfers data on the activity of a given unit towards another unit. It is also used in planning and modelling of experiments.

•mathematical modelling considers the different states of the unit and the way of their achievement. The unit is described by mathematical formulas.

Statistical regression models are most widely used in forecasting. The functional relationship between variables described by them may be linear or nonlinear.

Key moments in developing a regression model are: establishing regression coefficients, the choice of significant independent variables, and the verification of the adequacy of the model.

In the mid-1990s, a class of algorithms, known as ARIMA [1], were developed for time series forecasting on the basis of information contained in the previous history of the time series. These algorithms do not imply the use of a strictly defined forecasting model. The most appropriate forecasting method is selected on the basis of previous history and current values of indicators.

II. Neural Networks in Forecasting

Special attention should be paid to the option of applying neural networks for the purposes of forecasting. Their main advantage is the possibility of learning and improving forecasting precision. Neural networks represent statistical models based on analogy with the human brain. They are not substantially different in their methods and application from standard statistical models. They are used as statistical instruments in many fields: economics, psychology, statistics, etc.

A. Application of neural networks

Neural networks are used mainly to simulate a mathematical function with known input and output data. This is essential when the data complexity makes it very difficult or impossible to infer the function which generates them by mathematical methods. The tasks resolved by neural networks usually fall under one of the following categories:

•approximation of a function, regression analysis, dynamic data series forecasting and modelling.

•classification, structural recognition, image recognition, decision-making.

•data processing, including data filtering, grouping, compression.

Neural networks are applied in control and identification systems, transport management, process control, management of robots, artificial intelligence in computer games, decision-making, image recognition, radar systems management, identification of persons and objects, speech recognition, handwriting recognition, financial applications, data or knowledge accumulation.

B. Strengths

•Universal character - neural networks do not depend on the characteristics of input data, because there are no requirements for a specific input data distribution or for linearity of the function generating them.

•Easy-to-apply – the use of neural networks does not require any special training, because it is not necessary for their implementation to comprehend the intrinsic mechanism of their action – unlike statistical methods, which require knowledge in the field of probability theory and mathematical statistics.

•Neural networks are capable of modelling relationships between a large number of variables.

C. Weaknesses

•developing the network architecture for a specific task is complicated, because there are no standard patterns and it is necessary to construct the architecture starting from zero in each specific case.

•interpretation of the obtained results is difficult in some cases.

III. Application of the Balanced Scorecard in e-Learning for Evaluation of Higher Education

A Balanced Scorecard has been used for the evaluation of e-learning efficiency in higher education. It was created by Kaplan and Norton [2] in 1992 Since it was developed, the BSC has been implemented successfully in a number of business organizations and educational institutions. Some positive examples of this are:

• Edinburgh University [3]

• Texas Education Agency [4],

•University of Newcastle [5].

•Purdue University [6].

•Quality-driven Universities [7]

•[8] The authors compare the traditional methods of linear regression and neural networks for the forecast of costs in education. They apply the neural network methods for forecasting per pupil expenditures in the United States, comparing a neural network with backpropagation of the error signal and generalized regression with a multivariate regression model developed by the National Center for Education Statistics. They observed that the forecast is of much higher precision, if a simple linear neural network model is used than with the use of the statistical model.

•[9] The authors propose a methodology of evaluating technical education quality on the basis of the neural network approach. They establish common minimum quality indicator values, which should be reached in order for all stakeholders to be satisfied. They also developed special software for the forecasting of education quality from the point of view of the different participants in the education process.

•[10] The authors consider the issue of assessment of the quality of education of graduating students based on neural networks. They consider the importance of assessment as a main guarantee of quality in education and propose a non-linear assessment method using neural networks as a tool of non-linear programming.

•[11] According to the authors, the use of regression analysis to forecast success or failure of an individual student is not very efficient. For this reason, they use for this purpose the non-parametric statistical approach of neural networks. They establish that neural networks performance is not weaker than traditional methods and it is worth to explore their application.

The present study introduces an exemplar set of Key performance indicators for the evaluation of e-learning efficiency in higher education using a Balanced Scorecard /Table I/. For convenience, the data for the different years are normalized.

A complex approach was used in the evaluation of the knowledge and skills acquired by the students, applying a set of indicators to determine the level of student achievements.

•average marks at students' graduation. This indicator is used to ensure comparability of students' grades from different years, as well as from different fields of study and universities.

TABLE I
KEY PERFORMANCE INDICATORS OF E-LEARNING.

Pers-pective	Indicator	2001 x_1	2002 x_2	2003 x_3	2004 x_4	2005 x_5	2006 x_6	2007 x_7	2008 x_8	2009 x_9
Financial	Income from tuition fees	0,0000	0,0370	0,0741	0,1185	0,2074	0,3259	0,7407	0,9630	1,0000
	Income from scientific projects	0,0000	0,0377	0,0377	0,1506	0,1807	0,3313	0,6928	1,0000	0,9940
	Income from sponsorship	0,0000	0,1429	0,2857	0,3333	0,3333	0,2857	0,3810	0,6905	1,0000
	Income from sales of electronic study aids	0,0000	0,0541	0,0541	0,2162	0,2703	0,4595	0,6486	0,9730	1,0000
	Income from developing e-courses	0,0000	0,2857	0,2143	0,2857	0,5714	0,7143	0,8571	1,0000	0,4286
	Cost of amortization of tangible fixed assets and intangible assets	0,0000	0,0870	0,1304	0,4783	0,6957	0,8870	0,9130	0,9565	1,0000
	Staff salaries and social security costs of	0,0000	0,2727	0,3394	0,7273	0,9394	0,9636	1,0000	0,9212	0,8788
Customer	Number of applications per one student admission	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,5000	0,5000	0,0000
	Average entry marks of M.A. applicants	0,6000	0,4167	0,3667	0,1167	0,0000	0,4167	0,6167	1,0000	0,8333
	Rate of students satisfied with their studies %	0,2727	0,0000	0,3636	0,0000	0,6364	0,8182	0,9091	0,9091	1,0000
	Rate of students with tuition fees/stipends paid by business companies %	0,0000	0,7500	0,5000	0,5000	0,5000	0,7500	1,0000	1,0000	0,7500
Education process	Average time spent by a student to learn for one study course /h/	0,0000	0,6296	0,7407	0,9259	1,0000	0,9259	0,8889	0,8148	0,8519
	Average number of in-term tests (prior to the final exam) /number/	0,0000	0,2500	0,2500	0,5000	0,7500	0,7500	0,7500	1,0000	1,0000
	Degree of interactivity of study courses	0,0000	0,0000	0,0000	0,0000	0,0000	0,5000	0,5000	1,0000	1,0000
	Average time for F2F teaching per study course/h/	0,0000	0,1250	0,1667	0,2500	0,4583	0,6667	0,7500	0,8750	1,0000
	Students per teacher/number/	0,0000	0,1944	0,2500	0,3611	0,4722	0,6667	0,8333	0,9444	1,0000
	Number of new specialties	1,0000	0,0000	0,5000	0,5000	1,0000	0,5000	0,5000	0,5000	0,5000
	Number of new e-learning courses	0,1818	0,0000	0,3636	0,4545	0,6364	0,3636	0,4545	1,0000	0,8182
	Number of updated e-learning courses	0,6667	0,3333	0,0000	0,0000	0,3333	0,3333	0,6667	0,8333	1,0000
Staff	Percentage of teachers satisfied by the education process %	0,1818	0,0000	0,1818	0,3636	0,6364	0,1818	0,2727	1,0000	0,5455
	Number of publications per teacher	0,0000	0,5000	0,0000	0,5000	0,5000	1,0000	1,0000	0,5000	0,5000
	Number of attended conferences per teacher	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	1,0000	0,0000	0,0000
	Number of teachers with PhD degrees	0,5000	0,5000	1,0000	0,5000	0,0000	0,5000	0,0000	0,5000	1,0000
	Number of teachers having academic rank	0,0000	0,0000	0,3333	0,3333	0,6667	1,0000	0,3333	0,3333	0,3333
	Number of conferences on e-learning organized by universities	1,0000	1,0000	1,0000	1,0000	0,5000	0,5000	0,5000	0,0000	0,0000

•student success coefficient showing the average ratio of those who graduated successfully to the number of newly enrolled students. A high rate of student drop-out is characteristic for e-learning. Higher values of the success coefficient are directly associated with higher efficiency of education.

•a coefficient showing the ratio of students going into PhD study to the total number of graduates.

• a ratio of graduates working within their field of study to the total number of students graduated from e-learning.

•a ratio of the average salary of graduates working within their field of study to the average salary in the country.

By using these indicators, it is possible to simultaneously account for the students success rate, their competitiveness on the labour market, and the results from e-learning.

Our purpose is to use a neural network in order to find a functional relationship in a vector form between the indicators from Table I and Table II. Therefore, we will demonstrate that the neural network can utilize this functional relationship to make a forecast of the resulting values of indicators for 2009 – y9 on the basis of data from previous years. Table I contains

TABLE II
E-LEARNING EFFICIENCY INDICATORS

Y value	2001 y_1	2002 y_2	2003 y_3	2004 y_4	2005 y_5	2006 y_6	2007 y_7	2008 y_8	2009 y_9
Average graduation marks	0,1666	0,0000	0,1666	0,5000	0,8333	1,0000	0,6667	0,8333	1,0000
Student success coefficient	0,1176	0,0000	0,1176	0,4117	0,7058	1,0000	0,4117	0,7058	0,8235
Ratio of PhD students to graduates	0,0000	0,2500	0,5000	0,7500	0,7500	0,7500	0,7500	1,0000	1,0000
Proportion of graduates working within their field of study	0,0740	0,0000	0,0740	0,2592	0,4444	0,8148	1,0000	0,8148	0,8888
Average salary ratio of graduates working within their field of study compared to the average salary in the country	0,0000	0,1333	0,1333	0,4000	0,9333	0,9333	0,8000	0,9333	1,0000

the vector function arguments and Table II contains the results.
The state of BSC and e-learning efficiency indicators are described by and vectors. The vector function can be written in the following form:

$f(\vec{x}_n) = \vec{y}_n$,where
$\vec{x}_n = (\vec{x}_1, \vec{x}_2, ..., \vec{x}_9)$, $\vec{y}_n = (\vec{y}_1, \vec{y}_2, ..., \vec{y}_9)$,

The vectors $x_n = \vec{x}_1, \vec{x}_2, ..., \vec{x}_9$ are the values of key indicators for the evaluation of BSC measured on the relative scale [0, 1]

The vectors $y_n = \vec{y}_1, \vec{y}_2, ..., \vec{y}_9$ are the values of e-learning efficiency indicators measured on the relative scale [0, 1].

The neural network approach was used to determine the functional relationship between $\vec{x}$ and $\vec{y}$ vectors.

IV. The Experiment

The existence of a stable relationship between resulting e-learning efficiency indicators and BSC key indicators has been already demonstrated in practice. A common weakness of the abovementioned forecasting methods of mathematical statistics is their limitation by requirements regarding the characteristics of time series, so they often do not allow forecasting the real course of economic processes. Therefore, we propose the use of the neural network methodology in forecasts, in order to solve the problem of demonstrating a functional relationship between BSC key indicators and e-learning efficiency indicators.

To draw up a forecast of e-learning efficiency for 2009, we have data available from the 2001 – 2009 period. The task is, by using an appropriate neural network structure, to forecast the 2009 values of the efficiency indicators on the basis of the values of key BSC indicators for 2009 and of their previous history for the 2001-2008 period. For this specific task, the choice of the type and complexity of neural network depends on:

- the complexity of the defined task;
- how much data is available on education;
- the necessary number of entrances to and exits from the network.

A conclusion based on previous studies is that it is not absolutely necessary to use a neural network with radial basis function. The values of input indicators depend on a number of external factors like inflation, extensive growth of the educational institution, etc. The multidimensional space of available data is weakly covered by reference points. The radial basis function reduces to 0 the influence of indicators outside the limits of indicator reference values. Therefore, better results can be achieved using a perceptron with sigmoid activation nonlinear function of neurons.

During the experiment, it is necessary to check if there is a model with minimum possible complexity, as the data are scarce and not representative. Such a model is linear regression – a classic statistical method. As a minimal neural network, we can select a perceptron without hidden layer and with linear activation function of neurons. The most complex neural network model is the multilayer perceptron with 1 hidden layer of 10 neurons and a sigmoid activation function.
If this perceptron did not have nonlinearity, it would be analogous to a perceptron without a hidden layer, as a linear combination of linear combinations remains a linear combination. For this reason, we chose a neural network with a hidden layer with nonlinear activation function and we are looking for a nonlinear regression functional relationship between the different pieces of data. According to [12] the number of the neurons in the hidden layer coluld be estimated with :

$$\frac{mN}{1+\log_2 N} \le L_w \le m\left(\frac{N}{m}+1\right)(n+m+1)+m$$

where n – dimensionality of the input signal, L_w – needed amount of synaptic connections, m – dimensionality of the output signal, N – number elements in the training set. According to the same source, the number of neurons in the hidden layer /L/ could be estimated with :

$$L = \frac{L_w}{n+m}, \quad 2(L+n+m) \le N \le 10(L+n+m) \text{ and}$$

$$\frac{N}{10} - n - m \le L \le \frac{N}{2} - n - m .$$

The increase in the number of hidden layers in the multilayer perceptron is not necessary. It is proven that a multilayer perceptron with 1 hidden layer has universal approximation capabilities [13], [14], [15].

By testing a neural network on the training sample, we can determine the precision of reproduction of training examples, but we cannot estimate the precision of interpolation and/or extrapolation.

Design of the Experiment

The experiment was carried out through neural network training using the method of leave-one-out testing. The necessity of it came from the low volume of data that made it impossible to isolate a sample for evaluation of the error occurring in forecasting. In neural network training through leave-one-out testing, data on one of the years in the explored period is eliminated. Neural network training is carried out using the rest of the data and the error is determined as a difference between the forecast for the missing year and the available data for the year. This is repeated by excluding each of the years one by one in order to calculate an average value of the resulting error on all the training data set. Thus, the derived error is indeed an average value of the error in forecasting of the results for a forthcoming year.

The experiment was carried out in the following steps:

- Checking the condition for sufficiency of one-layer linear neural network in forecasting the indicator values. For this purpose, we train in a standard way a one-layer linear neural network and test it successively by training examples. The average error obtained in this way is 0.01, which shows that a one-layer linear neural network is fully sufficient for the purpose of forecasting.

•Using the method of training by leave-one-out testing, the average error in the forecast of one-layer linear neural network is determined. The average error in this case is higher to the point of being practically unacceptable – an average error of 0.4 can be compared to the width of the interval of indicator values. Only the fifth indicator is estimated properly. Even when the experiment is repeated several times using training by leave-one-out testing, the average and maximum error values stay at similar high levels.

•The same method is applied to determine the error of the forecast using a multilayer perceptron with 1 hidden layer of 10 neurons. The precision of the forecast is lower than the one of the linear neural network. It is not necessary to try to determine the precise number of neurons in the hidden layer, because the linear network deals much better with the task to be accomplished.

In our work on the project, we used the method of neural network pruning in order to improve forecast precision.

•We determined the error in training by leave-one-out testing of a pruned one-layer linear neural network. After pruning, the neural network is capable of reproducing the data on its training with an error within the limits of 0.01, part of the input indicators being eliminated – the ones that do not influence the output signal. The forecast precision of this pruned neural network increased so much that it became practically acceptable. Predictions of the error for the 2nd and 5th indicator differ by a few percentage points – about 10% of the exact width of the interval [0, 1], i.e. the forecast for these two indicators is within the range of real values, and not of the limits of the interval, as it was observed in the initial neural network and the multilayer perceptron. The precision of the forecast of the indicators is higher than for the neural networks before their pruning.

•The error is determined in training through leave-one-out testing of a pruned multilevel perceptron with one hidden layer of 10 neurons. In this case, the error is within 0.8, which is not an acceptable level.

•To develop a forecast, we use a pruned linear one-layer neural network (Fig. 5). The values of five indicators are forecasted for 2009 in 25 trials and at each trial, the neural network is trained again – the data are listed in Table IV, where the average values of the forecasted indicators are presented, after a correction by a confidence interval and the variance of forecast values (Table IV)

The following conclusions can be made on the basis of data in Table IV.

•the average forecast values are sufficiently close to the theoretically possible normalized values of indicators.

•irrespectively of the non-representative sample, a small deviation between the forecasts from the different trials is observed. The repeated training of the pruned network with

TABLE III
AVERAGE ERROR OF THE NEURAL NETWORK IN DIFFERENT CONFIGURATIONS AND TRAINING BY LEAVE-ONE-OUT TESTING

ANN topology	**Average graduation marks**	**Student success coefficient**	**Ratio of PhD students to graduates**	**Proportion of graduates working within their field of study**	**Average salary ratio of graduates working within their field of study compared to the average salary in the country**
Single layer	0,3366	0,3216	0,1583	0,2547	0,3959
Multilayer	0,3444	0,4819	0,2731	0,2710	0,4420
Single layer after pruning	0,1629	0,1619	0,1060	0,1504	0,2693
Multilayer after pruning	0,4273	0,4742	0,2178	0,228	0,3253

TABLE IV
RESULTS FROM 25 FORECASTING EXPERIMENTS

	Average graduation marks	**Student success coefficient**	**Ratio of PhD students to graduates**	**Proportion of graduates working within their field of study**	**Average salary ratio of graduates working within their field of study compared to the average salary in the country**
Average forecast	0,9421	0,7519	0,9197	0,8360	0,9824
Dispersion of predicted values	0,0386	0,1075	0,0804	0,0905	0,0425

new initial values of the weights of the synapses results in stable construction of relatively similar relationships between the output and the input signals of the network.

V. Conclusions Regarding the Results from the Experiment

•Using BSC allows to evaluate the efficiency of e-learning in higher education. The proposed set of indicators for evaluation of e-learning efficiency is appropriate for making forecasts of the development of this process in the future

•To make a forecast, it is not suitable to establish a nonlinear relationship between the initial and forecast values of the indicators by applying a multilayer perceptron with neurons in the hidden layer, as the examined period is too short.

•During the work process, alternative neural network structures were examined, the pruned linear network, in particular. The precision of the forecast was improved, reaching practically acceptable error levels.

•The applied method of pruning a neural network - Optimal brain surgeon – is a useful tool, increasing the probability of building networks with optimal architecture.

•The flexibility of neural network technologies presenting vast opportunities of choice of number of layers and neurons in the network, eliminating redundant neurons and synapses, and education on the basis of the algorithm of backpropagation of the error signal allows the construction of linear regression using a neural network without any hidden layer.

•Acceptable precision of the forecast of indicator values is achieved by linear neural network pruning. Using a hidden layer of the neural network turns out to be unnecessary.

•The decrease in the number of input signals in a pruned neural network can improve the precision of the forecast and make it less labour-intensive.

•The forecast of the indicator values for 2009 demonstrates stability of the results.

•On the graph of the structure of the pruned network, we can try to understand how the neural network “thinks” and how it uses the output indicators for the forecast. The comparison of the neural network pattern of “reasoning” with the theoretical and empirical knowledge can provide incentives for researchers in developing alternative forecasting methods.

Acknowledgment

This work is supported by the European Social Fund and Bulgarian Ministry of Education, Youth and Science under Operative Program “Human Resources Development”, Grant BG051PO001-3.3.04/40.

References

[1] Box, GEP and Jenkins, Time series Analysis: Forecasting and Control, GM 1976
[2] Robert S Kaplan, David P. Norton, Strategic Maps, 2006
[3] http://www.planning.ed.ac.uk/Strategic_Planning/BSC/Overview.htm
[4] Lauren Keller Johnson, Texas Education Agency: Boosting Performance and Accountability with the BSC, May 15, 2003
[5] Yee-Ching Lilian Chan, St. Thomas University:Which Balanced Scorecard to Use?”, McMaster University, AP Vol. 6 No. 4 — PC vol. 6, no 4 (2007) pages 399–414 © CAAA/ACPCdoi:10.1506/ap.6.4.4
[6] http://www.purdue.edu/ssta/general/strategicplan/bsc/files/BSC SSTA.pdf
[7] Elsa Cardoso, Maria José Trigueiros, Patricia Narciso, A Balanced Scorecard Approach for Strategy- and Quality-driven Universities
[8] Bruce D. Baker, Craig E. Richards, A comparison of conventional linear regression methods and neural networks for forecasting educational spending
[9] S.S. Mahapatra and M.S. Khan, A neural network approach for assessing quality in technical education: an empirical study
[10] HE De zhong,LIU Jing nan,ZHANG Su he, Evaluation of Academic Degrees and Graduate Education Based on Neural Network
[11] Bill C. Hardgrave1, Rick L. Wilson2, Kent A. Walstrom, Predicting graduate student success: A comparison of neural networks and traditional techniques
[12] V.V. Kuruglov, M.I. Dli, R. Golunov – Fuzzy logic and artificial neural networks, 2001, p. 59
[13] G. Cybenko Approximation by superposition of a sigmodial function, 1989
[14] Kurt Hornik, Maxwell Stinchcombe, Halbert White, Multilayer feedforward networks are universal approximators, 1989
[15] Ken-Ichi Funahashi, On the approximate realization of continuous mappings by neural networks, 1989, *to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.