Prediction of e-Learning Efficiency by Neural Networks

Petar Halachev

Institute of Information and Communication Technologies, 1113 Sofia
Email: halachev@iit.bas.bg

Abstract: *A model for prediction of the outcome indicators of e-Learning, based on Balanced ScoreCard (BSC) by Neural Networks (NN) is proposed. In the development of NN models the problem of a small sample size of the data arises. In order to reduce the number of variables and increase the examples of the training sample, preprocessing of the data with the help of the methods Interpolation and Principal Component Analysis (PCA) is performed. A method for optimizing the structure of the neural network is applied over linear and nonlinear neural network architectures. The highest accuracy of prognosis is obtained applying the method of Optimal Brain Damage (OBD) over the nonlinear neural network. The efficiency and applicability of the method suggested is proved by numerical experiments on the basis of real data.*

Keywords: *e-Learning efficiency, Balanced ScoreCard, Neural Networks.*

1. Introduction

Recently the use of e-Learning in higher education is increasingly expanding. The assessment of its efficiency allows the identification of the gaps of high schools activities and the measures for their reduction. This contributes to the increase of their efficiency and competitiveness on the market of educational services.

On the other hand, it is also necessary to evaluate the final results of e-Learning, namely the graduation marks of the students, their competitiveness on the labour market and their motivation for further training.

For prediction of the outcome efficiency indicators of e-Learning on the basis of data from BSC, a Neural Network (NN) method is chosen.

NN are becoming still more widely used for analysis, modeling and predicting, in solving specific problems in management processes – risk assessment, forecasting of stock indices, control of operations with credit cards, forecasting of exchange rates, prices of securities, etc. [1, 2, 3]. The prognoses in these areas are characterized by a high level of errors. The contemporary theoretical and practical achievements in the area of neural networks allow the development of predicting models with high accuracy.

2. Predicting the efficiency of e-Learning by NN

The neural network method was applied to predict various aspects of e-Learning – efficiency, probability of successful completion of the training course, quality, satisfaction of the participants in training, etc.

L y k o u r e n t z o u, G i a n n o u k o s, M p a r d i s, N i k o l o p o u l o s and L o u m o s [4] recognize that the growing use of e-Learning requires the establishment of a mechanism for predicting the marks of the students at an early stage of the course. The authors apply a method for predicting the marks of the students in an online course lasting 10 weeks. In the method proposed a multilayer neural network without a feedback is used, in order to predict the marks of the students. The students are divided into two groups according to their graduation marks. The results of the multiple choice tests were used as input data for the model. The obtained results show that the creation of an accurate prognosis is possible after the third week of the 10-weeks course. The low levels of the false prognosis demonstrate the adequacy of the chosen method. The prognosis by NN is compared to a prognosis made by linear regression applied to the same problem. The conclusion is that NN are more efficient at all stages of forecasting. The proposed methodology can assist the instructors in providing better educational services and indicate which students need specific support.

In order to predict the indicator "students, that have dropped out of e-Learning", L y k o u r e n t z o u, G i a n n o u k o s a, N i k o l o p o u l o s a, M p a r d i s a and L o u m o s a [5] develop a method based on three different types of neural network architectures. The architectures of NN which are used are as follows: a neural network without a feedback: Support Vector Machine, and SFAM (Simplified Fuzzy ArtMap (Adaptive Resonance Theory Mapping)) by K a s u b a [6]. The authors note that a NN architecture may fail to accurately classify some of the students, and offer three different schemes of taking decision, which combine the results of three NN architectures. They compare the results obtained by other methods used on similar problems and find that the solution obtained by NN is significantly better.

B a k e r and R i c h a r d s [7] predict the costs for the education of the students in USA by NN and compare them with a prognosis prepared by a multivariate regression model, developed by the National Center for Educational Statistics. They found out that the prognosis is more accurate when prepared by a linear neural network in comparison to the statistic model.

Z h o n g, H e and N a n [8] propose a nonlinear estimation method and use NN as a tool for nonlinear modeling to assess the quality of education of the graduating students.

3. Assessment of the efficiency of e-Learning by Balanced Scorecard

R o b e r t K a p l a n and D a v i d N o r t o n [9] have created a strategic model for assessment of the efficiency named Balanced SCorecard (BSC).

The application of BSC by the educational institutions starts at a later stage, compared with the use of BSC in business organizations. However, universities in USA, Great Britain, Australia and Russia implement successfully BSC for assessment and management of their activities [10-13].

The basic components of a BSC model are the Key Performance Indicators (KPI). These are qualitative and quantitative indicators, which serve to assess the efficiency of the activity of the organization and to show the degree of achievement of its tactical and strategic goals. The key indicators are used also to assess the difficulty to measure the activities, such as satisfaction of the concerned parties, staff motivation, competitiveness of the offered service, etc. A set of KPI for assessment of the efficiency of e-Learning at a Bulgarian university with e-Learning is presented in Table 1. The data is from a master course for the period 2001-2009.

Table 1. KPI and perspectives

Perspective		Key Performance Indicator			
Perspective	**Financial**	Income from tuition fees (Euro)	Number of applications per one student admission	**Users**	**Perspective**
		Income from scientific projects (Euro)*	Average entry marks of M.A. applicants		
		Income from sponsorship (Euro)	Rate of students satisfied with their studies (%)		
		Income from sales of electronic study aids (Euro)*	Rate of students with tuition fees/stipends paid by business companies (%)*		
		Outcome for developing e-courses (Euro)			
		Cost of amortization of tangible fixed assets and intangible assets (Euro)*			
		Staff salaries and social security costs (Euro)			
	Educational process	Average time spent by a student to learn for one study course (h)	Percentage of teachers satisfied by the education process (%)	**Staff**	
		Average number of in-term tests, prior to the final exam (number)*	Number of publications per teacher		
		Degree of interactivity of study courses*	Number of attended conferences per teacher		
		Average time for F2F teaching per study course (h)*	Number of teachers with PhD degrees		
		Students per teacher (number)*	Number of teachers having academic rank		
		Number of new specialties	Number of conferences on e-Learning organized by universities*		
		Number of new e-Learning courses*			
		Number of updated e-Learning courses			

Except KPI, for comprehensive assessment of the efficiency of e-Learning, some outcome indicators are required, which reflect the results of the learning process: grades of the students, their competitiveness on the labour market, their motivation for further training and relative percentage of workers in the specialty. To predict the values of the outcome indicators on the basis of the values of KPI (Table 1), a comprehensive neural network model is offered in the present paper.

The selected e-Learning efficiency indicators are:
- average graduation marks;
- ratio of graduated students to enrolled students;
- ratio of PhD students to graduates;
- proportion of graduates working within their field of study.

4. Neural network model for predicting the efficiency of e-Learning

The purpose of the present paper is to build an adequate and accurate model for practical application in forecasting the efficiency of e-Learning by the tools of NN. The training by NN is carried out by submitting KPI at its input (Table 1) and at the output of NN – the outcome indicators that measure the efficiency of e-Learning for the period 2001-2008. Then, by submitting at NN input KPI only for 2009, the values of the outcome efficiency indicators are predicted. The outcome indicators obtained at the NN output are compared with their actual values for 2009, and thus the accuracy of NN prognosis is evaluated.

4.1. Data needed for the creation of a NN model

The sample of the actual data is gathered for the period 2001-2009 at a Bulgarian university that applies e-Learning.

The sample data has the following characteristics:
- the number of the examples is less than the number of variables,
- the investigated period of time is relatively short.

NN in conjunction with Principal Component Analysis (PCA) are used in order to solve the above mentioned problems.

According to B r a c e [14], "The number of tested examples has to exceed the number of predictor variables", i.e., for application of Principal Component Analysis (PCA) more examples than the number of variables are needed. In this case the available data are for 9 years (9 examples), 25 input and 4 outcome variables. This requires interpolation of the data.

4.2. Stages in the development of a NN model

The development of a predictive NN model comprises the following stages:
- centering and normalization of the data;
- correlation analysis of the data. At this step the indicators which have big correlation coefficients between them are removed (these are the indicators, marked by the symbol * in the third and fourth column of Table 1);

- interpolation of the data using the Hermite method;
- correlation analysis of the data in order to examine to what extent the correlation coefficients between the separate indicators are preserved after the interpolation;
- principal Component Analysis of the KPI;
- development of NN. During the development the steps are as follows:
- NN is trained by data;
- a prognosis for the values of the outcome indicators for 2009 is prepared;
- the error is calculated.

Step 6 is repeated with the following NN architectures:
- linear NN is developed;
- linear NN is developed with Optimal Brain Damage;
- nonlinear NN is developed;
- nonlinear NN is developed with Optimal Brain Damage.
NN architecture with the smallest prognosis error is selected as appropriate.

4.3. Construction of a predictive model

4.3.1. Data centering and normalization

The application of PCA requires that the data it is applied on must be centered and normalized [15]. For centering of a variable a standard formula is used, by determining its arithmetic average and for normalization – a formula by its dispersion.

4.3.2. Correlation analysis of the data in order to reduce the dimension

When performing PCA it is necessary to analyze the correlation coefficients between the variables and to eliminate those variables which correlation with other variables is greater than 0.9. During the statistical processing of data when two variables have a correlation coefficient greater than 0.9, it is assumed that they measure the same indicator and only one of them must remain in the dataset. In Table 1 the symbol * marks the variables which are removed from the dataset due to their high correlation.

The correlation analysis of the output indicators shows that the correlation coefficients between the e-Learning efficiency indicators do not exceed 0.85 and therefore no removal is required.

4.3.3. Data interpolation

According to the type of function $f(x)$ in the sections between the interpolation nodes, the type of interpolation can be linear, parabolic, bilinear, bi-cubic, and others, depending on the chosen function. Bi-cubic spline in the form of Hermite is a suitable function, since the interpolates will not exceed the maximum value of the reference data, nor descend below their minimum value (such as the "graduate mark" key indicator).

The interpolation of the input and output data is preceded by determination of the necessary number of interpolators. Typical for the interpolation is that in the data noise, which increases in direct proportion to the increase of the number of interpolants "is introduced".

In order to reduce the noise to minimum levels, it is necessary to select a minimum number of points, sufficient for PCA. The requirement towards the data is that the number of examples must be greater than the number of variables. The accomplishment of the necessary calculations shows that for interpolation of the variables it is sufficient to create one intermediate point for each data point. After this step the sample contains 17 examples (9 points, 8 intermediate intervals with one interpolated point in each of them).

4.3.4. Correlation data analysis to verify the result of the interpolation

To verify to what extent the data dependencies are preserved, it is necessary to make correlation analysis of the input and output variables before and after interpolation, and to compare the results. If the data dependencies are preserved, the difference in correlation coefficients before and after interpolation is minimal. The average change of the correlation coefficients after interpolation of the input indicators is within the range of 0.01 - 0.03%, and that of the outcome indicators is zero.

It can be concluded that the correlations between the variables before and after interpolation is preserved within normal limits.

4.3.5. Principal Component Analysis

The method Principal Component Analysis is offered by Pearson and serves to reduce the dimensionality of the data.

The principal components (results of PCA) are new independent variables, describing the whole sample. From a mathematical point of view PCA is an orthogonal linear transformation that transforms the data into a new coordinate system so that the projection of data with greatest variance lies on the first coordinate axis (called the first principle component), data with second largest variance lie on the second coordinate axis, etc. [17].

After conducting PCA using statistical software, 4 principal components are obtained, which together reflect 90.239 % of the variance in original data.

5. Numerical experiments through simulation of NN in order to predict the efficiency of e-Learning

NN simulation proceeds as follows:
- NN is trained with data by the method Leave one out [18];
- prognosis of the values of the outcome indicators for 2009 is prepared;
- the error is calculated.

5.1. Linear neural network simulation

The outcomes of training a linear NN with data for 2001-2008 and its testing by passing the input data for 2009, are presented in Table 2. The average error from a linear NN is 13.39 % of the total width of the intervals of the outcome variables. From a statistical point of view it seems too large; to improve the accuracy of the prognosis a method for optimization of NN is applied.

The optimization of a linear NN is done by the algorithm OBD (Optimal Brain Damage) [19]. OBD is a method for optimizing the structure of the NN that works by removing unimportant neural connections. By removing the unimportant connections from a network, several improvements can be expected: better generalization and fewer training examples required. The basic idea is to use second-derivate information to reduce the network complexity and the training set error. The neural networks structure that is obtained is presented in Fig. 1, where the synaptic weights that drop out of the NN structure are marked by dotted lines:
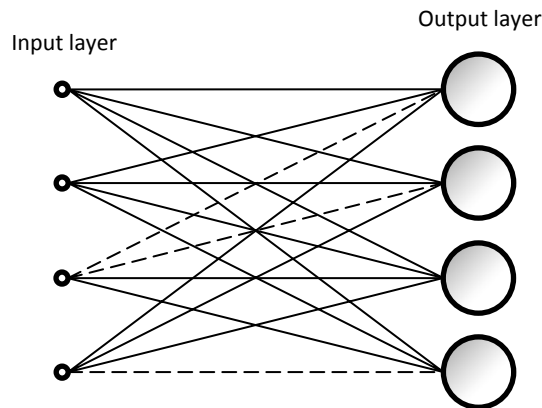

Fig. 1. A scheme of the optimized linear network

The average outcomes are presented in Table 2.

The error of 6.63 % shows improvement of the predicting capabilities of the optimized NN compared to the non-optimized.

- Simulation of the nonlinear network.

It is necessary to determine the number of hidden layers and the number of neurons in each of them in a nonlinear network.

In terms of the calculation capabilities only one hidden layer is sufficient. It is proven that a nonlinear neural network with one hidden layer has universal approximation capabilities [20-22].

The references research reveals that no rule for determining the number of the neurons in the hidden layers exists [23]. In order to calculate the number of neurons K r u g l o v at al. [24] proposes a formula. The calculations with the help of this formula show that five is a reasonable number for the neurons in the hidden layer. Therefore a neural network with five neurons in the hidden layer is selected.

The activation function that is used by the neurons in the hidden layer is sigmoidal. The graph of this function is in the shape of letter "S". This is the most

commonly used function to find nonlinear dependencies in the design of neural networks. This function is strictly increasing in the interval [-∞; +∞] and its output is limited to [−0; 1].

The function is given by:

(1)
$$\varphi(v) = \frac{1}{1 + \exp(-v)}.$$

Its name "logistic" function comes from Pierre François Verhulst who used it to study the rate of growth of populations back in 1844. The graph of this function is given on Fig. 2.
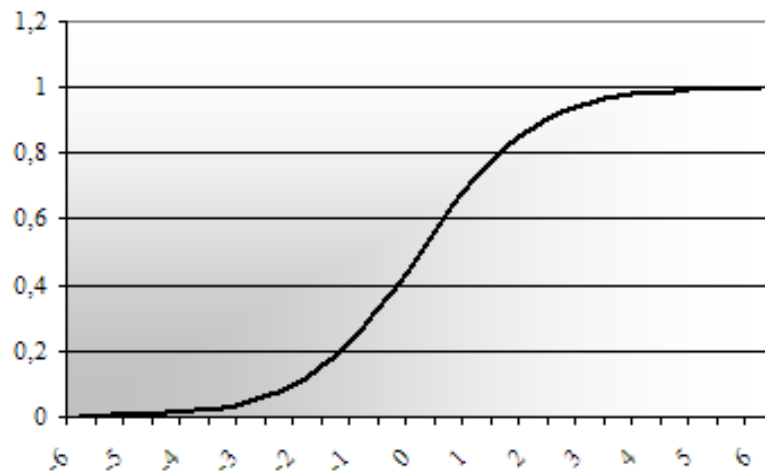

Fig. 2. Graph of the sigmoidal function

The activation function used in the output layer is linear.

The results of the simulation of a nonlinear NN built with these considerations in mind are presented in Table 2.

The average prognosis error amounts to 15.06 % of the total width of the interval of outcome indicators.

The method OBD is applied on the nonlinear NN. The NN obtained as a result is presented in Fig. 3. The dotted lines denote the synaptic weights dropped out by OBD method.

The average result of the training of an optimized nonlinear neural network is presented in Table 2.

From a statistical point of view the error rate is within the tolerance of 5 %. The accuracy of the prognosis of an optimized nonlinear NN is higher than that of a linear NN. The error in the prognosis, expressed as a percentage of the width of the interval of the efficient variables is 3.71 % (Table 2). This is the lowest level of error and seems satisfactory from a statistical point of view.
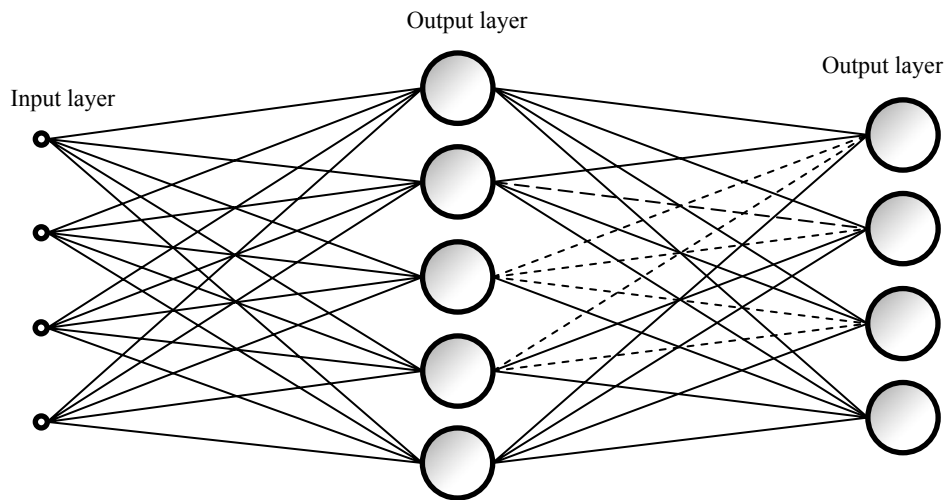
Fig. 3. Scheme of the optimized nonlinear network

Table 2. Results of the experiment

Type	Indicator								Average error
	Average graduation marks		Ratio of graduated students to enrolled students		Ratio of PhD students to graduates		Proportion of graduates working within their field of study		
	Value	Error	Value	Error	Value	Error	Value	Error	
Width of interval (min-max)	0.49		0.15		4.00		21.00		
Target value (2009)	5.08		0.83		1.00		58.00		
Linear NW	5.15	14.29	0.83	0.00	2.00	25.00	61.00	14.29	13.39
Linear NW with OBD	5.14	12.24	0.83	0.00	1.00	0.00	61.00	14.29	6.63
Nonlinear NW	5.15	14.29	0.82	6.67	2.00	25.00	61.00	14.29	15.06
Nonlinear NW with OBD	5.12	8.16	0.82	6.67	1.00	0.00	58.00	0.00	3.71

The data in Table 2 averages the simulations carried out for every NN architecture. Retraining of the NN with other initial values of the weights of remaining in it synapses, leads to the construction of relationship between the input and output signals of the network. The values obtained from prognosis of the indicators reflecting the efficiency of e-Learning for 2009 are close to the real ones.

6. Conclusion

A model of a neural network for obtaining accurate predicting structure with small data samples is proposed in the present paper. The data is processed in advance by correlation analysis, interpolation and PCA. In this case the outcome indicators for e-Learning efficiency are predicted on the basis of key performance indicators from BSC.

The proper selection of the number of neurons in the hidden layer and the optimization algorithm maximize the prognosis accuracy. The nonlinear NN reaches acceptable prognosis accuracy only after optimization. Reducing the number of synaptic weights in the neural networks is useful in practice, because it increases the prognosis accuracy.

The results of prediction by a nonlinear network are closer to the real values, compared with the results of prediction by a linear NN. Therefore a conclusion could be made for the presence of nonlinear relationships in the data.

The performed numerical experiments on the basis of real data show an acceptable from a practical point of view (3-4 %) prognosis error. Thus the efficiency and the practical applicability of the proposed mathematical model for predicting the outcome indicators of e-Learning efficiency on the base of neural networks and PCA, are proved.

This paper does not claim to discuss and resolve the whole spectrum of problems in assessing and predicting the efficiency of e-Learning in higher education. Its purpose is to give some recommendations and conclusions related to the various aspects of e-Learning evaluation. In such context this paper could be seen as a step towards our further studies in the area of the topical problem.

R e f e r e n c e s

1. J o h n, M. Economic Forecasting Challenges and Neural Network Solutions. Computer Science Dept., Oregon Graduate Institut, USA, 1995.
2. H o u, Z a i-E n, F u-J i a n D u a n. The Neural Network Method of Economy Forecasting. World Congress on Software Engineering, IEEE, 2009.
3. L e o n a r d o, M. S., R. B a l l i n i. Design a Neural Network for Time Series Financial Forecasting: Accuracy and Robustness Analisys. Instituto de Economia, Universidade Estadual de Campinas, Sao Paulo-Brasil, 2008.
4. L y k o u r e n t z o u, I., I. G i a n n o u k o s, G. M p a r d i s, V. N i k o l o p o u l o s, V. L o u m o s. Early and Dynamic Student Achievement Prediction in e-Learning Courses Using Neural Networks. DOI: 10.1002/asi.20970, 2008.
5. L y k o u r e n t z o u, I., I. G i a n n o u k o s V. N i k o l o p o u l o s,, G. M p a r d i s, V. L o u m o s. Dropout Prediction in e-Learning Courses through the Combination of Machine Learning Techniques. School of Electrical and Computer Engineering, National Technical University of Athens, Zographou Campus, 15773 Athens, Greece, 2009.
6. K a s u b a, A. EXPERT. Simplified Fuzzy ARTMAPT. Miller Freeman, Inc., 1993.
7. B a k e r, B. D., C. E. R i c h a r d s. A Comparison of Conventional Linear Regression Methods and Neural Networks for Forecasting Educational Spending. 1999.
8. Z h o n g, H e D e, L i u J i n g N a n, Z h a n g S u H e. Evaluation of Academic Degrees and Graduate Education Based on Neural Network. – Journal of Chongqing University (Natural Science Edition), 2003.
9. K a p l a n, R., D. N o r t o n. The Balanced Scorecard: Translating Strategy into Action. 2006.

10. J o h n s o n, K. J. Boosting Performance and Accountability with the BSC. Texas Education Agency, 2003.
11. L i l i a n, C. Y.-C. Which Balanced Scorecard to Use. St. Thomas University, 2007.
12. Purdue University, SSTS: Strategic Plan - Balanced Scorecard. 2009.
13. C a r d o s o, M. J. T. P. N. E. A Balanced Scorecard Approach for Strategy-and Quality-Driven Universities. 2005.
14. N i c o l a, B. R. K. R. S. SPSS for Psychologists. Published by Palgrave Macmillan, Australia, 2006.
15. V u n d e v. Records on Applied Statistics, **Vol. 2**, 2003.
16. E d w a r d, J. J. A User's Guide to Principal Components, 2004.
17. S u i k o v a , S. T. I. Statistical Study. Publishing House "Luren", 2000.
18. S. M. Cross-Validatory Choice and Assessment of Statistical Predictions. 1974.
19. C u n, Y a n L e, J. D e n k e r, S. S o l l a. Optimal Brain Damage. Advances in Neural Information Processing Systems, 1990.
20. C y b e n k o, G. Approximation by Superposition of a Sigmodial Function. 1986.
21. F u n a h a s h i, K.-I. On the Approximate Realization of Continuous Mappings by Neural Networks. 1989.
22. H a r t m a n, K., K o w a l s k i. Layered Neural Networks with Gaussian Hidden Units as Universal Approximations. 1990.
23. H a y k i n, S. Neural Networks: A Comprehensive Foundation. 1994.
24. K u r u g l o v, V. V., M. I. D l i, R. G o l u n o v. Fuzzy Logic and Artificial Neural Networks. 2001.