U.P.B. Sci. Bull., Series C, Vol. 72, Iss. 4, 2010 ISSN 1454-234x

# USING ARTIFICIAL NEURAL NETWORKS IN E-LEARNING SYSTEMS

Elena ŞUŞNEA[1]

*În această lucrare propun un mod de utilizare a rețelelor neuronale artificiale de tip multi-layer percepton (MLP) şi radial basis function (RBF) prin modelarea unui predictor al performanțelor studenților înscrişi la un curs desfăşurat în sistem e-learning. În scopul măririi vitezei de calcul, am optimizat arhitecturile rețelelor neuronale prin micşorarea numărului de neuroni ce compun rețeaua respectivă. Rezultatele pe care le-am obținut confirmă performanțele acestor tipuri de predictori, ratele de eroare obținute fiind foarte scăzute.*

*In this paper I present a method that uses artificial neural network, multi-layer percepton (MLP) and radial basis function (RBF), through modeling a performance predictor of the students attending an e-learning course. In order to increase computation speed, I optimized neural networks architecture through decreasing the number of neurons used in the respective network. The results I achieved sustain the performance of those types of predictors, the error rates gathered being very low.*

**Keywords:** artificial neural networks, data mining, classification, ILIAS

## 1. Introduction

In the present, the e-learning systems allow acquisition and storage of a huge volume of data. Adequately exploited, this data allow gathering of information that can lead to a good knowledge of the process developed within the educational institutions [1].

Diversification of the software technologies had a concrete result the solutions regarding the management of different and complex activities implied by the e-learning system. Learning Management Systems (LMS) are IT web-based applications that present the educational content at the right time and in the appropriate format assuring the interaction between the student and the learning resources [2]. In such an educational system which is student-centered, the data regarding the activities carried out by the students throughout the entire course, if correctly interpreted, can provide a useful feedback for the personnel responsible with the proper progress of the teaching activities (managers, professors, software developers, specialists in psycho-pedagogy).

[1] Lecturer, National Defence University "Carol I", Bucharest, Romania, e-mail: esusnea@yahoo.com

The LMS system that will be referred to throughout this paper is ILIAS [3]. This allows the efficient fulfillment of integral courses or materials for courses by means of some instruments and standard models for the working and learning processes in the e-learning system. Also, ILIAS provides efficient administration and tracking of the following types of resources:

- Educational resources – the system makes available a series of resources stored on the server, including Web pages, which present the educational contents, the homework and the periodical and final exams, the links to other educational resources available on web, simulations etc;
- Data referring to the users' activities such as: creating, modifying, assessing or using the educational resources, the log-in and log-out moments.

Among the factors that influence the increase of data quantity stored in ILIAS system, we mention: existence of some courses in progress implies data storage for each student regarding accessibility to the course content, all the assessment tests, storage of the answers and the number of attempts made for each item etc.; appearance of new courses implies not only storage of previously mentioned data but also creation of new educational resources according to the student's profile; updating of the educational contents, by replacing the parts considered "old-fashioned" with others that are in accordance with some recent scientific results and can be better applied; diversification the electronic format regarding presentation of the educational materials by replacing some static formats with others that have bigger degree of interaction and are more suggestive.

Taking into account the facts that have been previously presented, we intend to solve an aspect regarding extraction of some knowledge from the data sets of the students that had on-line training courses by using the educational resources made available by ILIAS e-learning platform. In order to extract new information that is useful for knowledge of the educational processes, we will use the data mining classification techniques [4], as well as the neural networks and the *k*-nearest neighbor.

## 2. Using artificial neural networks within the e-learning educational systems

It was impossible for the appearance and development of the new techniques not to have a strong impact upon the military system, too [5]. Development of the military processes and of these phenomena is tightly connected to the technological development. To professionalize the military personnel implies, among others, using an advanced system of instruments and teaching technologies including advanced distributed learning at distance [6].

The data base that is the object of the present study contains data collected throughout a period of two years, taken from the tests that have been given to the military and civilian students. The courses which represent the assessment objective are military [7].

The students have been assessed online with a questionnaire containing 25 questions, realized according to the teaching principles with a standard degree of question difficulty. Each assessment was registered at the end in a unique database.

ILIAS system presents statistic data regarding the assessment activity for each student during the period of an online course. Fig.1 shows a part of the statistic data achieved at a student's online assessment activity, made available for the teachers by the ILIAS platform.

Detailed Evaluation for

| Statistical Data | |
|---|---|
| Test Results in Points | 24 of 25 (96.00 %) |
| Test Results in Marks | passed |
| Questions already worked through | 25 of 25 (100.00 %) |
| Number of Test Passes | 1 |
| Scored Pass | 1 |
| Time of Work | 00:24:32 |
| Average Time of Work | 00:00:49 |
| First Visit | 2008-12-08 19:56:00 |
| Last Visit | 2008-12-08 21:46:33 |
| Mark of Median | passed |
| Rank of Participant | 13 |
| Rank of Median | 15 |
| Total Number of Participants | 29 |
| Median of Test Result in Points | 24 |

| Question Results for Pass 1 | | |
|---|---|---|
| 1. | Q9 | 1 of 1 (100.00 %) |
| 2. | Q8 | 1 of 1 (100.00 %) |
| 3. | Q62 | 4 of 4 (100.00 %) |
| 4. | Q58 | 1 of 1 (100.00 %) |
| 5. | Q5 | 1 of 1 (100.00 %) |

Fig. 1. Example of presentation with the detailed results

The students have the right to try to answer twice for each question, the final answer having two possibilities: correct or wrong (incorrect) answer. The maximum score that can be obtained at this test is of 25 points, for the 25 questions, all the questions having equal values. It has been registered for each question the number of attempts, the rightness of the answer as well as the time necessary to give the answer.

Next we are going to study the means of adjusting the data existing in the database offered by ILIAS system, containing answers to the questionnaires given by 476 students in order to extract the information using a prediction model. A series of classifying parametric items will be presented in order to compare the performances on the real database of the online assessing system. This study allows observing the way in which the data mining techniques can be applied for identification “valid, novel, potentially useful, and understandable patterns” in database [8]. The error rates are used to compare the performances of each implemented classifying indicator.

The main problem is to discover a set of features in order to be able to classify the students. In this situation we will be able to build a predictor that would shape each student after giving the online test. With this instrument it is possible that a student be better helped in order to efficiently use the system resources. The most difficult phase is the one of efficient pre-processing of the data for classifying.

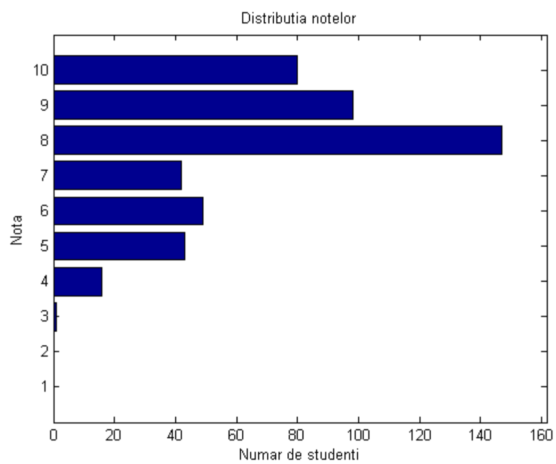


Fig. 2. Scoring distribution for the 476 students

***The data set and the class labels***

The first step in creating the predictor is extracting from the database the samples and data that are to be processed. The chosen database has 476 students that have answers for each unique set of 25 questions. The results achieved by them are presented in Fig.2.

The students can be grouped according to their score in a number of classes, as it follows:

- 10 classes that can be identified as being those associated to the grade as a whole, between 1 and 10 (table 1);

- 3 classes that can be identified as being associated to students *poor, good* or *very good* (table 2).
- 2 classes that can be identified as being associated to the students that have graduated the course, respectively those who have not graduated (table 3).

*Table 1*

**Classifying the students in 10 classes**

| Class label (grade) | Number of students | Percent [%] |
|---|---|---|
| 1 | 0 | 0 |
| 2 | 0 | 0 |
| 3 | 1 | 0,210084 |
| 4 | 16 | 3,361345 |
| 5 | 43 | 9,033613 |
| 6 | 49 | 10,29412 |
| 7 | 42 | 8,823529 |
| 8 | 147 | 30,88235 |
| 9 | 98 | 20,58824 |
| 10 | 80 | 16,80672 |

*Table 2*

**Classifying the students in 3 classes**

| Class label | Number of students | Percent [%] |
|---|---|---|
| 1 | 60 | 12,60504 |
| 2 | 238 | 50 |
| 3 | 178 | 37,39496 |

*Table 3*

**Classifying the students in 2 classes**

| Class label | Number of students | Percent [%] |
|---|---|---|
| 1 | 17 | 3,571429 |
| 2 | 459 | 96,42857 |

The classifying process is based on a set of features such as:

1. Success rate (total number of correct answers);
2. Number of correct answers reached at first attempt;
3. Number of necessary attempts in order to reach the correct answer;
4. The necessary time to give a correct answer to the questions;
5. The necessary time to give either a correct or incorrect answer to the questions.

***The classifying indicators***

The data mining techniques are used in a series of applications from different fields [9] , including academic education [10]. As a consequence, it is impossible to create a unique classifying indicator that would allow getting the best results regardless the application. The best classifying indicator in each situation depends on the problem area. In practice, we can also have a situation in which no classifying indication can offer acceptable solutions, and in this case it is useful to separate them according to their specialties and to use the decisional fusion techniques (combining the classifying indicators).

The most popular classifying indicators that are used are MLP (multi-layered perceptron) and RBF (networks with radial based functions) [11]. These are used in the practical problems of classification. In order to be used in the most efficient way, the database needs to be pre-processed by means of a reduction algorithm for the dimension, such as PCA (the analysis in main components). In the end, the classifying indicators are assessed according to the achieved error rate.

For a practical work, that data must be first grouped, namely reduced to the variation interval of [-1, 1] and zero average, according to the statistical values of each of the five characteristics, in order of this we are using the formula: $x_i = (x_i - \mu)/\sigma$, where $\mu$ and $\sigma$ represent the average and the standard deviation of the initial set. This is correct only if starting from the *a priori* hypothesis that data have Gaussian distribution.

By grouping the data, each set of features has a normal distribution with zero average and unitary standard deviation. Moreover, the classifying indicators impose the additional requirement that all data should have the variation field between the same limits.

In case the data are divided in three classes, the difference between the error rates achieved by processed data and by unprocessed data is obvious (table 4). Processed data allow an error much more inferior to the one achieved for the unprocessed data. The way to achieve the prototypes of the two neural networks will be presented in the following section.

*Table 4*

**The results achieved by the two classifying indicators for processed and unprocessed data**

| The classifying indicator | Error rate with processed data | Error rate with unprocessed data |
|---|---|---|
| MLP | 0,326 | 0,989 |
| RBF | 0,354 | 1,012 |

***The database for training and testing. Validation of the classifying indicators.***

For the training and testing of the classifying indicators, the data must be divided in distinct data bases, as it follows: for the training of the classifying indicator, with a part of the main database a new database is created which is called training database, and for testing the results of the classifying indicators, the rest of the unused database is created in the testing database. The way in which the database is divided is important, as it represents an improvement criterion for the classifying indicator [12].

In the present situation, the method of subdividing the *k-fold* database has been used. This implies dividing the database in approximately equal $k$ subsets. Training of the classifying indicator will be done $k$ times, each time leaving one subset for testing. In a limit case, if $k$ is equal to the number of students, the method is also known as *Leave One Out*.

The bigger the $k$, the better the classifying result. To test this hypothesis it has been used a validation of the classifying indicator with $k$=2 and with $k$=7. The data selected to create the two subsets are randomly chosen and independent. In the case of $k$=2, the database is divided in two subsets equal in dimensions (238 students in the training database and 238 students for the testing database). In the case of $k$=7, the database is divided in 7 subsets (with 68students for each), every subset being used in turn as a testing database. As a consequence, the training base will be 85.71% from the database, and the testing/ checking base will be 14.29% from the database.

Watching the results achieved in table 5, it can be noticed that for a bigger $k$, the error rate is smaller. Thus, for the rest of the study, the technique of validating the classifying indicators with $k$=7 will be used.

*Table 5*

**The results achieved with the three classifying indicators for the *k-Fold* technique with *k*=2 and *k*=7**

| The classifying indicator | Average error rate for $k$=7 | The error rate for $k$=2 |
|---|---|---|
| MLP | 0,326 | 0,357 |
| RBF | 0,354 | 0,417 |

***Improving the classifying indicators***

In the case of the two neural classifying indicators, there is a series of ways to choose the architecture. The architecture is created starting from the number of features for the input and the number of classes where the database is distributed. If the number of features is equal to 5 (there are not used techniques for dimension reduction), both neural networks will have the input layer composed of 5 neurons. To simplify the problem, the students are divided in the three classes, according the grade. Thus, the neural networks will have output layer composed of three neurons. In conclusion, the parameter that needs to be selected for architecture improvement is the number of neurons of the layer hidden between the input layer and the output. Selection of the best number of neurons is done as it follows:

1) It starts from a hidden layer which is composed of a number of neurons equal to the biggest layer of the network (in the present situation, the input layer , equal to 5 neurons) and the performances of the classifying indicators are assessed (the error rate);
2) Gradually a neuron is added to the hidden layer and the performances of the classifying indicators are re-assessed;
3) The optimum number of neurons of the hidden layer is achieved when the error rate is minimum or has no significant fluctuations.

Network training is done by reaching a stopping criterion represented by the maximum number of iterations (epochs=300) or a minimum error value (0,100).

In the tables 6, 7 and 8 there are presented the results for improvement of the two neural networks (MLP and RBF) for different class divisions. Inside the tables, the configurations with the optimal number of neurons the hidden layer have been marked in bold. In the case of an error rate which is identical for several configurations, the one having the smallest number of neurons in the hidden layer was considered optimal.

*Table 6*

**Error rates of the neural networks (the case with three classes and five loggings)**

| Classifying indicators | Number of neurons in the hidden layer | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| MLP | 0,411 | 0,398 | 0,382 | 0,377 | 0,357 | 0,356 | **0,355** | 0,359 | 0,361 |
| RBF | 0,398 | 0,392 | 0,385 | 0,361 | 0,355 | **0,350** | 0,350 | 0,349 | 0,349 |

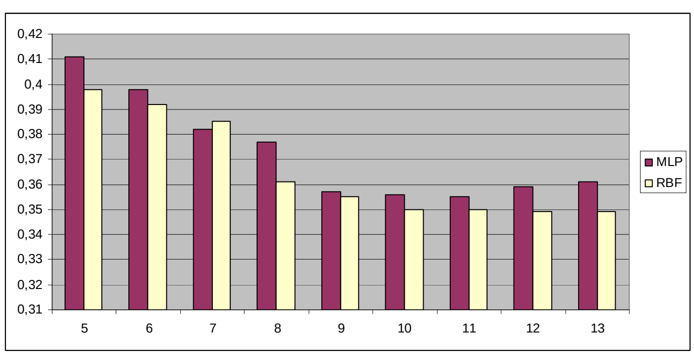


Fig. 3. Error rates of the neural networks (the case with three classes and five inputs)

*Table 7*

**The error rates of the neural networks (the case with ten classes and five inputs)**

| The classifying indicator | The number of neurons in the hidden layer | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| MLP | 0,678 | 0,669 | 0,650 | 0,647 | **0,646** | 0,646 | 0,647 | 0,648 | 0,646 |
| RBF | 0,598 | 0,595 | 0,588 | 0,587 | 0,587 | **0,586** | 0,587 | 0,588 | 0,590 |

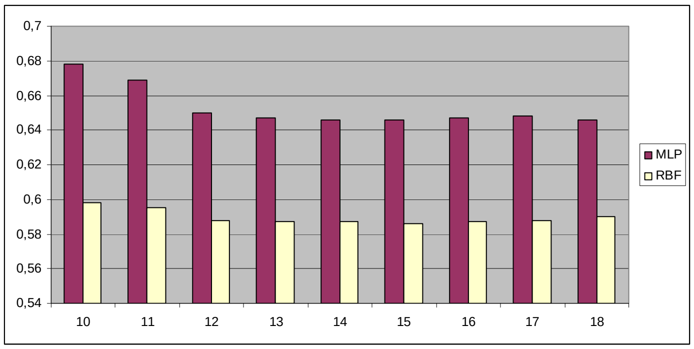


Fig. 4. The error rate of the neural networks (the case with ten classes and five inputs)

*Table 8*

**The error rate of the neural networks (the case with two classes and five inputs)**

| The classifying indicator | Number of neurons in the hidden layer | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| MLP | 0,311 | 0,386 | 0,323 | **0,211** | 0,211 | 0,210 | 0,210 | 0,209 | 0,209 |
| RBF | 0,194 | 0,186 | 0,173 | 0,172 | **0,166** | 0,167 | 0,167 | 0,166 | 0,165 |

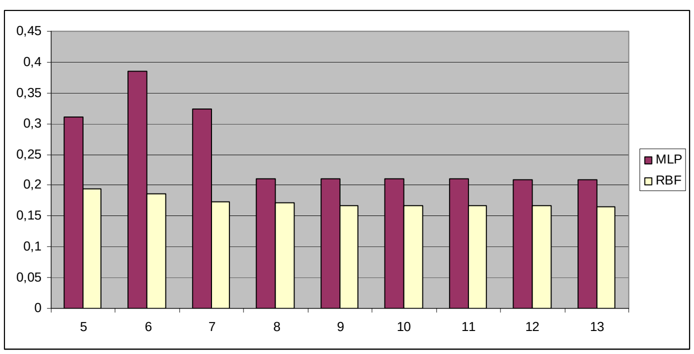


Fig. 5. The error rate of the neural networks (the case with two classes and five inputs)

The smaller the error, the closer to perfect the generated model is [11]. This allows us to estimate better the performances of the students that appear after the configuration.

## 3. Conclusions

From the result analysis I made, it can be noticed that big error rates are achieved in the case of a greater number of classes. This is justified by the fact that the problem the neural network intends to solve is much more complex. This implies training a network whose architecture would allow discrimination in the data hyperspace of some regions with different shapes and dimensions.

We can also notice that the RBF network has in most situations error rates inferior to MLP network. This leads to the conclusion that the data follows a multidimensional Gaussian distribution which can be much easier discriminated by means of the functions with radial base from RBF network components.

By using the neural networks we can have the best configuration from the error point of view. They also allow improvement of the performances along with the growing of the database by retraining, thus including the new information collected from the subsequent students. This shows the dynamic character of the model.

## R E F E R E N C E S

[1] *O. B. Alaba, I. Făgărăşan, R. Dobrescu, S. St. Iliescu*, System analysis for e-learning grids, UPB. Sci. Bull., Series C, **Vol. 71**, Iss. 3, p. 71-78, 2009
[2] *P. Rogers, G. Berg, J. Boettcher, C. Howard, L. Justice, K. Schenk*, Enciclopedia of distance learning 2nd ed., Information Science Reference, New York, 2009
[3] www.ilias.de
[4] *E. Şuşnea*, Using data mining techniques in higher education, in The 4th International Conference on Virtual Learning, p. 373, 2009, [www.icvl.eu/2009]
[5] *I. Roceanu, V. Popescu*, Metrics and requierements in Learning Management System, in The 4th International Conference on Virtual Learning, p. 80, 2009, [www.icvl.eu/2009]
[6] *Advanced Distributed Learning*, Case study: U.S. Navy Apprentice Technical Training School, Alexandria, 2009 [http://www.adlnet.gov]
[7] http://adl.unap.ro
[8] *U.M. Fayyad, G. Pitatesky-Sapiro, P. Smyth, R. Uthurasamy*, Advanced in Knowledge Discovery and Data Mining, in AAAI/MIT Press, Menlo Park, CA., p. 5, 1996
[9] *E. Şuşnea*, Classification techniques used in Educational System, in The 4th International Conference on Virtual Learning, , p. 377, 2009, [www.icvl.eu/2009]
[10] *M.F. Matei,* ANOVA in the educational process, U. P. B. Sci. Bull., Series C, **Vol. 70**, No. 3, p. 121-129, 2008
[11] *C. Molder*, Recunoaşterea formelor, vol. 2 Metode de clasificare, Editura Academiei Tehnice Militare, Bucureşti, 2004
[12] *M. Panait, A. Morega*, Neural and nervous reflex networks: a new approach to integrated control, UPB. Sci. Bull., Series C, **Vol. 71**, Iss. 3, p. 159-171, 2009.