FORECASTING THE SUCCESS OF EDUCATION SEEKERS FROM A SEPARATE EDUCATIONAL COMPONENT BASED ON THE RESULTS OF THE PRELIMINARY MASTERY OF SUBJECT COMPETENCIES

The paper examines the main concepts related to the quality of education in general and the assimilation of educational material by higher education seekers. The task of predicting a seeker's grade in any discipline is formulated with data on his assimilation of program learning outcomes that also correspond to this discipline. The available specialized information system of own development is described which applies a number of methods (multivariate linear regression, artificial neural networks, k-nearest neighbors) and determines which method will be the most effective for the analysis of specific data. It is noted that during the further improvement of the quality system of knowledge assessment, it is important to determine at what level the student of education possesses the acquired competences, i.e. to calculate the success of seekers in terms of general and professional competences and program learning outcomes, determined by the standards of higher education and educational programs developed on their basis. The developed algorithm for calculating the success rate of higher education applicants in terms of program learning outcomes is presented; according to this algorithm, data were prepared on the acquisition of software creation competencies by 78 seekers of the first level of higher education of the educational and professional program Intelligent Decision Support Systems specialty 124, Systems analysis, of the DSEA. To solve the problem of forecasting by the method of artificial neural networks, the programming and data analysis language R is proposed. A script for finding the optimal neural network architecture is created. It was found that the best result (correlation is 0.9599, average absolute reduced error is 0.1132, percentage of correctly predicted points on the Ukrainian scale is 79.2%) provides a perceptron with two hidden layers and five neurons in each one. This network was applied to predict the success of the new academic group: correlation is 0.923, the average absolute reduced error is 0.0654, the percentage of correctly predicted points on the Ukrainian scale is 82.4. The obtained results can be used to assess the quality of the structural and logical scheme of the EPP and in the work of the department during the analysis of seekers’ success, etc.

Keywords: educational and professional program, forecasting, artificial neural network, perceptron, neural network training, R-language.
level of assimilation of new knowledge by an individual depends primarily on his diligence and the basic level of knowledge, which is almost constant during the study period, therefore a sharp deviation in the evaluation process may indicate the presence of problems, objective and subjective factors that influence on the educational process. All educational components at the corresponding level of higher education, the list of academic disciplines and the logical sequence of their study, the number of ECTS credits, as well as the expected learning outcomes and competences that must be mastered by the seeker of the corresponding level of higher education, are contained in the educational program [2]. The structural and logical scheme of training is provided in the form of a network of interdisciplinary connections and is valid throughout the entire period of implementation of the corresponding training program. So, one of the factors affecting the grade received by a higher education degree seeker from a separate discipline are grades from the disciplines that "support" it, i.e. those that precede this one.

**Analysis of the subject area**

In paper [3], to improve the management of the educational process at the graduation department of a higher education institution, it was proposed to develop a specialized information system which applies a number of methods (multivariate linear regression, artificial neural networks, k-nearest neighbors) and defines a method that will be the most effective for the analysis of specific data. The task was formulated as follows. A seeker's grade in any discipline needs to be predicted, using given data on grades in "supporting" disciplines:

\[
y = \{x_1, x_2, ..., x_n\},
\]

where \(y\) is the predicted grade in the discipline;
\(x_i\) is a grade in the \(i\)-th "supporting" discipline;
\(n\) is the number of "supporting" disciplines.

The researcher sets the limit values of the parameters for each method (for example, the minimum and maximum number of hidden layers of artificial neural networks, etc.), then the software system performs the calculation for each of the methods, the results are added to the table. The researcher chooses the best, in his opinion, method (usually based on the minimum total error, but other selection criteria are also possible). The selected method is used to predict the grades of the same subject for a new group. The choice of supporting disciplines is not clearly a point of the given algorithm, since it is carried out either directly from the educational program (its structural and logical scheme), or as a result of some previous research.

The operation of the system was described using the example of data on higher education seekers majoring in System Analysis [2] and the subject of programming knowledge and skills acquisition; four input and one output factors were identified:

- \(x_1\) is a grade in the discipline Programming and Algorithmic Languages;
- \(x_2\) is a grade in the discipline Algorithms and Data Structures;
- \(x_3\) is a grade in the discipline Mathematical Logic and Theory of Algorithms;
- \(x_4\) is a grade in the discipline Optimization Methods and Operations Research;
- \(y\) is a grade in the discipline Technology of Creating Software Products.

Names of groups and surnames of seekers are informative factors.

Conducted research with the help of the developed system proved that, from a number of mathematical methods (linear regression analysis, artificial neural networks, nearest neighbors), the method of artificial neural networks leads to satisfactory results of predicting the grades of higher education seekers from a separate discipline, depending on the quality of assimilation of the previous material. By conducting a series of numerical experiments, the optimal architecture of the neural network was selected, a two-hidden-layer perceptron with five neurons in each. The obtained results can be applied in the work of the department during the analysis of seekers' performance, etc.

The application of the method of artificial neural networks for solving the forecasting problem is also described in papers [4–7].

**Problem formulation and input factors**

However, in the further improvement of the quality system of knowledge assessment, it is important to determine at what level seeker of higher education possesses the acquired competences, i.e. to calculate the success of higher education seekers in terms of general and professional competences and program learning outcomes, determined by the standards of higher education [8] and educational programs developed on their basis [9]. A software system has been created that makes it possible to work with a list of formed competencies in subjects and program learning outcomes both within standards of higher education and EPP/ESP [10]. A feature of the developed system is the ability to analyze the success of the applicant or the entire group in terms of competencies (GC, PC) and program learning outcomes (PLO) that they have mastered. In work [11], the data analysis of the student of SA-20-mag group of the educational and professional program Intelligent Decision Support Systems of the second level of higher education, academic specialty 124, System Analysis, was carried out. Examples of calculating the success rate of the best student and the average indicator of competencies and program learning outcomes were presented.
The disadvantages of the system are the ability to work only within one academic group and the development of the curriculum in its entirety, taking into account all disciplines of free choice. And if the first drawback is overcome by downloading and storing the results of individual calculations for further processing, then taking into account the factor of "selective disciplines" has no solution since the list can change annually. In addition, a number of PLOs does not arouse interest from the point of view of the "professional image" of the graduate.

The following is suggested. First, only mandatory educational components will be considered. Secondly, an algorithm for calculating the success rate of higher education applicants in terms of program learning outcomes will be created [12].

Take the following notations:
- \( \text{OK} \) is a set of all mandatory components of the educational program;
- \( \{\text{OK}_{\text{select}}\} \subseteq \{\text{OK}\} \) are mandatory components that are considered;
- \( \{\text{PLO}\} \) is a set of program learning outcomes of an educational program;
- \( \{\text{PLO}_{\text{select}}\} \subseteq \{\text{PLO}\} \) are program learning outcomes that are considered;
- \( N \) is the number of mandatory components of the educational program;
- \( M \) is the number of program learning outcomes of the educational program;
- \( C_r \) is the amount of credits assigned by the educational program for the mandatory component \( \text{OK}_i, i = 1..N; \)
- \( K_i \) is the number of program learning outcomes, the mastery of which is provided by the mandatory component \( \text{OK}_i, i = 1..N; \)
- \( \text{Cr}_K = C_r / K_i \) is the number of credits for one program learning outcome, mastery of which is provided by the mandatory component \( \text{OK}_i \in \{\text{OK}_{\text{select}}\}, i = 1..N; \)
- \( Z \) is the number of seekers who received grades based on the results of the final control of mandatory components;
- \( B_j \) is the grade received by the \( j \)th higher education seeker according to the results of the final control from the mandatory component \( \text{OK}_i \in \{\text{OK}_{\text{select}}\}, i = 1..N, j = 1..Z; \)
- \( \text{NB}_j = B_j \cdot \text{Cr}_K \) is a grade recalculated per share of one program learning outcome;
- \( \text{NOK}_k \) is the number of program learning outcomes, the mastery of which is provided by the mandatory component \( \text{PLO}_k \in \{\text{PLO}_{\text{select}}\}, k = 1..M. \)

The calculation table is filled with data according to the formula:

\[
R(\text{PLO}_k) = \sum_{i=1}^{N} \text{NB}_{ji} / Z, k = 1..M, \text{OK}_i \in \{\text{OK}_{\text{select}}\}, \text{PLO}_k \in \{\text{PLO}_{\text{select}}\}.
\]  

(2)

Next, the data is normalized, i.e. brought to the accepted 100-point scale, after which various actions are possible. For example, determining the best achievers according to individual PLOs, comparing the learning results of different academic groups, carrying out clustering, i.e. grouping the achievers depending on the level of mastery of program learning outcomes, etc.

However, the main interest of research is in predicting a seeker's grade in any discipline based on data on his/her assimilation of the "supporting" program learning outcomes:

\[
y = \{\text{PLO}_1, \text{PLO}_2, ..., \text{PLO}_n\}.
\]  

(3)

**Data preparation and problem solving using artificial neural networks**

As in paper [3], grades of the 78 DSEA students from groups SM-13-1, SM-14-1, SM-15-1, SM-16-1 in Systems Analysis major are used (meaningful contents of the training courses and teachers have not changed in four years, the form of teaching has not changed either, there have been neither quarantines nor martial law).

According to [2], the optional discipline Technology of Creating Software Products (BK-2.7) provides improvement of knowledge according to three program learning outcomes:

- PLO 08. To have modern methods of developing programs and software complexes and making optimal decisions regarding the composition of software, algorithms of procedures and operations;
- PLO 09. To be able to create effective algorithms for computational tasks of system analysis and decision support systems;
- PLO 13. Design, implement, test, adopt, support, operate software tools for working with data and knowledge in computer systems and networks.

Next, information on all subjects is summarized in Table 1.
The method of artificial neural networks for solving the problem of forecasting. The programming and data analysis language R was used for calculations. This language is intended for statistical data processing and work with graphics. It is also a free and open-source programming environment developed within the framework of the GNU project. Available libraries allow to apply modern methods, including the method of artificial neural networks for solving the problem of forecasting.

A script has been created, the basic part of which is presented in the listing:

\[ N_{layer} = \left\lfloor \frac{N_{x} + N_{y} + 1}{N_{y}} \right\rfloor \]

Then the number of neurons \( N_{layer} \) in a two-layer network can be determined by the formula:

\[ N_{layer} = \left\lfloor \frac{N_{w}}{N_{x} + N_{y}} \right\rfloor \]

The number of neurons in the hidden layer can be specified in the process of setting up the neural network using a constructive algorithm [14]. Accordingly, the primary number of neurons is assumed to be equal to the minimum number (calculated by formula (5)). In case of unsuccessful training, one neuron is added to the hidden layer, the weight coefficients of which are assigned random values. The addition of neurons continues until the quality of the neural network reaches the required value.

The number of neurons in the hidden layer can be specified in the process of setting up the neural network using a constructive algorithm [14]. Accordingly, the primary number of neurons is assumed to be equal to the minimum number (calculated by formula (5)). In case of unsuccessful training, one neuron is added to the hidden layer, the weight coefficients of which are assigned random values. The addition of neurons continues until the quality of the neural network reaches the required value.

The programming and data analysis language R [15] was used for calculations. This language is intended for statistical data processing and work with graphics. It is also a free and open-source programming environment developed within the framework of the GNU project. Available libraries allow to apply modern methods, including the method of artificial neural networks for solving the problem of forecasting.

Next, a decision using the method of neural networks is made. The number of hidden layer neurons is related to the amount of training data and the required number of inputs and outputs of the network. The number of neurons in the hidden layers can be estimated using the inequality for estimating the number of weighting coefficients necessary for mastering a given number of examples in the training sample [13]:

\[ \frac{N_{w}N_{p}}{1 + \log_{2}N_{p}} \leq N_{w} \leq N_{y} \left( 1 + \frac{N_{w}}{N_{y}} \right), \left( N_{x} + N_{y} + 1 \right) + N_{y} \]
mydataframe <- read.table(paste(getwd(),"/StudentsC.txt",sep=""), header=TRUE, sep=";")

After numerous runs of this script for different parameters of the number of hidden layers and the number of neurons, it was found that the best result (correlation is 0.9599; average absolute reduced error equals 0.1132; percentage of correctly predicted points on the Ukrainian scale is 79.2) provides a perceptron with two hidden layers and five neurons in each one (Fig. 1, Fig. 2).
Next, the network is applied to a new academic group. Result: correlation equals 0.923, the average absolute reduced error is 0.0654, the percentage of correctly predicted points on the Ukrainian scale is 82.4 (Fig. 3).

Fig. 2. Calculation results

Next, the network is applied to a new academic group. Result: correlation equals 0.923, the average absolute reduced error is 0.0654, the percentage of correctly predicted points on the Ukrainian scale is 82.4 (Fig. 3).

Conclusions

The conducted studies proved that the application of the method of artificial neural networks for predicting the grades of higher education seekers in a separate discipline is leading to satisfactory results. By conducting a series of numerical experiments, the optimal architecture of the neural network was selected, a two-hidden-layer perceptron with five neurons in each.

The obtained results can be used to assess the quality of the structural and logical scheme of the EPP and in the work of the department during the analysis of seekers’ success, etc.

References

1. Regulations on the internal system of ensuring the quality of education. URL: http://www.dgma.donetsk.ua/docs/acts/Положення_про_внутр_сист_заб_якості_ДДМА_2020_Сайт.pdf (17.03.2024)
8. Approved standards of higher education. URL: https://mon.gov.ua/ua/osvita/visha

<table>
<thead>
<tr>
<th>Author</th>
<th>Position</th>
<th>Institution</th>
<th>E-mail</th>
<th>ORCID ID</th>
<th>ResearcherID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oleksandr MELNYKOV</td>
<td>PhD in Technics</td>
<td>Associate Professor of the Department of Intelligent Decision Making Systems, Donbas State Engineering Academy, Kramatorsk, Ukraine</td>
<td><a href="mailto:alexandr@melnikov.in.ua">alexandr@melnikov.in.ua</a></td>
<td>0000-0003-2701-8051</td>
<td>B-3142-2016</td>
</tr>
<tr>
<td>Veniamin GITIS</td>
<td>PhD in Technics</td>
<td>Associate Professor of the Department of Intelligent Decision Making Systems, Donbas State Engineering Academy, Kramatorsk, Ukraine</td>
<td><a href="mailto:veniamin.gitis@gmail.com">veniamin.gitis@gmail.com</a></td>
<td>0000-0002-7434-8259</td>
<td></td>
</tr>
<tr>
<td>Iryna GITIS</td>
<td>student in Kharkiv National University of Radio Electronics, Kharkiv, Ukraine</td>
<td>gitis. <a href="mailto:iryna@gmail.com">iryna@gmail.com</a></td>
<td></td>
<td>0009-0009-9091-8854</td>
<td></td>
</tr>
</tbody>
</table>