Development and Research of Algorithms for the Formation the Individual Educational Trajectories of Students in the Digital Educational Platform

Veronika V. Zaporozhko
Orenburg State University
Orenburg, Russia, 460018
zaporozhko_vv@mail.osu.ru

Denis I. Parfenov
Orenburg State University
Orenburg, Russia, 460018
fdot_it@mail.osu.ru

Maria Lapina
North Caucasus Federal University
Stavropol, Russia, 355017
mlapina@ncfu.ru

Daniele Sora
Sapienza University of Rome
Roma, Italy, 00185
sora@dis.uniroma1.it

Abstract

At present, the widespread use of modern information technologies in education, including open online courses, ensures the sustainable development of a single digital educational environment. However, one of the key problems of the mass approach to learning is the construction of individual educational trajectories. Taking into account the individual characteristics of each student is an urgent necessity. Achieving this goal is quite feasible when teaching students on individual learning routes. In this paper we have investigated two approaches to solving this problem. The first approach is based on the use of a genetic algorithm that allows you to form the optimal learning route, designed to meet the personal educational needs and individual capabilities of each student of the online course. The second approach involves the mathematical apparatus of neural networks give recommendations on the further optimal formation of an individual educational trajectory. The paper presents the results of experimental studies and examples of individual trajectories formed on the basis of the proposed algorithms.

1 Introduction

An important direction in improving the education system in many countries is the implementation of the concept of a modern digital educational environment. Massive open online courses (MOOCs) are considered as one of the main components of this environment. The advantage of MOOCs is that the learner gets access

In: Jože Rugelj, Maria Lapina (eds.): Proceedings of SLET-2019 – International Scientic Conference Innovative Approaches to the Application of Digital Technologies in Education and Research, Stavropol – Dombay, Russia, 20-23 May 2019, published at http://ceur-ws.org

to world-class knowledge regardless of his location, social status and other characteristics that are significant for traditional forms of education [Zap17]. Due to a significant number of students on MOOC platforms, it is possible to collect a large amount of data to form a profile of the learner. However, today the lack of an individual and differentiated approach to each student is one of the main drawbacks of MOOCs. This is due to the fact that existing MOOC platforms mainly implement only one pre-determined learning route by the course author. Therefore, one of the important areas for improving MOOCs is to provide complex personification. The need for personification of learning is a reflection of the natural desire for mankind to take an individual approach to personal opportunities, features, requests and preferences [Par18]. As part of this study, we proposed an approach that allows the intellectual management of individual educational trajectories in a particular online course. This approach is an algorithmic solution that forms the optimal trajectory of an individual learning route for each student by forming sets of learning objects. The basis for the algorithm for managing individual educational trajectories in MOOCs is the approach based on a hybrid combination of the Heuristic and cybernetic Data Mining methods. In particular, a genetic algorithm is used to form an individual set of learning objects. The constructed algorithm allows to dynamically development and correct the individual educational trajectory of each student depending on a set of parameters: diagnostic questionnaire results, tests score, features of perception and memorization of the material and others.

2 Related Work

At present, the amount of research devoted to the problem of development an individual educational trajectory in the implementation of the concept of the digital educational environment is permanently growing. Here are presented various approaches to the generation of individual learning route.

Researchers from the National Taiwan Normal University suggested using adaptive computer testing to identify problems in mastering individual blocks in the online course learning process. The database stores information about courses with given coefficients of difficulty. Based on the results of testing, the selection of appropriate courses with the lowest coefficient of labor input is carried out. Using the obtained data, the automated system generates an optimal individual training program for each student, using a genetic algorithm [Hon05].

A group of Taiwan scientists in their study suggested solving the problem of identifying the ability to learn and the difficulty level of the recommended curriculum's to each other. This problem is key when generate an individual learning route. To collect data within the framework of the study, the scientists conducted the assessment of students after mastering each block of educational content. The evaluation was carried out through computerized adaptive testing. The test results were then used to form the optimal route for each student. The approach proposed in the study is based on the hybrid use of the genetic algorithm and the case-based reasoning [Hua07].

Samia Azough et al. (Morocco) used a genetic algorithm to generate pedagogical paths which are adapted to the learner profile and to the current formation pedagogical objective. In its study they developed the description of an adaptive e-learning system. The system proposed by the authors allows the learner to study courses adapted to his profile. To implement adaptive learning, researchers applied two-step work of the genetic algorithm. At the first stage, the proposed mechanism is used to form optimal trajectories for the search for learning goals, taking into account data from the student's profile. At the second stage, the results obtained were adapted using data obtained from social networks [Han18].

A team of researchers from the University of Alcala (Spain) investigated how to perform dynamic selection of learning objects based on the genetic algorithm for constructing a course structure depending on the input set of competencies (formed in the learner) and the output (planned learning outcomes) [Mar11].

A number of studies of scientists from China are devoted to the preparation of individual tasks in the test form using genetic[Yao14]. The proposed approaches are summarized and implemented in the form of the Online Automatic Test System for various MOOC platforms.

A team of researchers from Russia proposed a method for constructing an individual learning route that meets the requirements of the user. In order to form learning paths, authors use domain ontology, on the basis of which separate learning objects are selected. Each learning object is complemented by sets of input and output competences that are ranked according to the Bloom's Revised Taxonomy (Remembering, Understanding, Applying, Analyzing, Evaluating, Creating). The genetic algorithm is used to construct the most appropriate educational trajectory from the available learning objects [Shm15].

Pragya Dwivedi et al. (India) formed individual educational trajectories in an online environment using a genetic algorithm with a variable length representation. The application of this algorithm provides a flexible duration of the recommended training course for each learner based on his learning style and level of knowledge. The original data is exported from the student's profile. Individual educational trajectories are built, taking into account information about the graduates of the course [Dwi05].

As part of his research, Sibel Somyürek (Turkey) provided an overview of adaptive multimedia education systems for the period from 2002 to 2012. The review presented by the author made it possible to identify technological trends and approaches in this area of research. The author made a special emphasis on modular structures, intellectual analysis of data, methods of machine learning and neural networks [Gol89].

Zacharoula Papamitsiou and Anastasios A. Economides (Macedonia) carried out research in the field of Learning Analytics and Educational Data Mining, revealing the impact of these technologies on adaptive learning [Bha10].

Researcher F. Okubo et. al. (Japan) suggested methods that predict student estimates using the Recurrent Neural Network based on journal data stored in educational systems [Bha10].

A team of researchers from Taiwan in their study conducted a definition of the learner's learning style based on his behavior in the browser. The proposed approach is based on the use of the multi-layer feed forward neural network (MLFF). In this study, the authors touched on several factors on the basis of which they assessed the behavior of the learner in the browser. In their opinion, the main factors are the following: the use of built-in auxiliary devices (ESD), navigation through links. The use of this approach has made it possible to adapt the educational environment to the needs and capabilities of the learner [Hua07].

In a study by R. Stathacopoulou et. al. (Greece) an approach is presented in which neuro-fuzzy synergism is used to evaluate students in the context of an intellectual learning system. In this study, the authors created a model of the student, on the basis of which he can evaluate information about his knowledge [Par18].

In the framework of the study by Cristina Conati et. al. (U.S.A.) in an intellectual learning system explored the capabilities of the Bayesian neural network for conducting a long-term assessment of knowledge, determining the plan and predicting the actions of students. The authors noted that the advantage of the proposed approach is the update of the network in real time. For this, an approximation algorithm based on a stochastic sample was used. Information from the model is used to help students and adapt the support system [Mar11].

Thus, the conducted review of researches has shown the urgency of development optimal individual learning routes and their correct in real time. At the same time, the heuristic algorithms are the main tool that allows the most effective management of individual educational trajectories.

3 Approaches For Forming Of An Individual Learning Trajectory In MOOC's

In this research to solve the problem of forming an individual learning trajectory in MOOC we will use a hybrid approach based on a combination of intelligent methods of data analysis and heuristic algorithms. When forming an individual trajectory, we will take into account the peculiarities of students' perception and memorization of information (in other words – from the learning style) [Azo10]. At the same time, using the neural networks algorithm the educational route will be formed on the basis of the learning objects included in the online course units. Let us dwell in more detail on the implementation of the neural networks approach.

Let us present the initial data for solving the claimed problem with the help of the mathematical tools of the genetic algorithm. Having analyzed the subject area of the task, we have identified the following tuple, characterizing the formation process of the individual educational trajectory (IET).

$$IET = (S, C, P), \tag{1}$$

where $S = (s_k)$ – the set of students learning a particular MOOC, k – number of students, $K \epsilon N$; $C = Unit_x$ – MOOC, located in a cloud-based learning environment and consisting of units, x – number of units in a particular course, $x \epsilon N$.

Each unit of the MOOC contains a specific set of content groups. Then let $G = g_1, \ldots, g_n$ – the set of generalized content type groups, when n – number of these groups, n=4. Each group contains a certain set of learning objects $g_i = LO_{i,j}$, where $LO_{i,j}$ – the set of LO_s in each unit, belonging to the selected generalized group g_i (Table **??**). $Unit_x = G_1, \ldots, G_4$. Then $P = P_1, \ldots, P_n$ is a valid set of individual routes for each student. Each individual learning route should consist of a specific set of $LO_{i,j}$ different types such as Presentations (slides), textbooks with diagrams, flowcharts, pictures, etc.); Infographics (mind maps, charts, diagrams, etc.), illustrations (pictures, posters); Webinars (video online meetings); Video lessons, recording screencasts, animated video clips (2D 3D animation); Audio conferencing and online meetings; Audio notes; Audio lessons (recordings); Workbooks audio; Glossaries (thesaurus, dictionaries); Reading (lecture notes, ebooks, tutorials, manuals, reports, articles, interactive textbooks, documents); Quizzes (or tests); Assignments (self-reports, tasks, essays, exercises, project works, mini action researches); Games (educational games, including simulation video games, virtual worlds); Virtual laboratories (interactive training systems); Interactive learning models; Workshops.

Each learning object $LO_{i,j}$ can take part in the formation of an individual learning route with its mandatory entry into a generalized group g_i. For the purposes of formalization, we introduce the Boolean variables 0 or 1, which describe alternatives to the selection of learning objects, i.e. $LO_{i,j} = 0, 1$.

Each object of the sets G and S can be represented as a set of attributes that numerically characterize these objects. Attributes are defined on a limited set of positive values. The definition of characteristics and values of attributes (parameters) for the identified sets. The task of determining the value of the attribute coefficient and the relative weight of the attribute is solved using empirical data, obtained as a result of the questionnaire, and expert estimates. To identify the relative weights of these attributes, experts were asked who ranked attribute values in descending order of importance.

The weight of each unit in the course is determined by the following formula:

$$W_{Unit_x} = \prod_{h=1}^{D} (\mu_{h,s_k})^{b_h}, \tag{2}$$

where μ_{h,s_k} - attribute coefficient value μ_h for unit depending on the particular type of student s_k, b_h – relative weight of attribute g_h for unit .

To select an individual learning route in MOOC, you also need to find the weight of the student. The weight of each student is determined by the following formula

$$W_{S_k} = \prod_{y=1}^{Z} (a_{y,s_k})^{\nu_h}, \tag{3}$$

where a_{y,s_k} - attribute coefficient value a y for student s_k, ν_h - relative weight of attribute a_y.

In the process of optimization under consideration, the parameter space under study is sufficiently large. The task does not require a strict global optimum, so it is sufficient to find an acceptable, most suitable (effective) solution in a short time.

To solve the optimization problem, we developed an algorithm for controlling the process of formation of an educational trajectory using a neural network. Compared to existing analogues, the algorithm uses heuristic analysis of data streams generated by students during the MOOC study. All the data collected are classified and ranked according to their importance, both for the particular course and for the learner. The flexibility of the proposed solution is due to the ability to dynamically change the educational trajectory in the course of the course study. The proposed solution is transparent for the trainee, forming an educational trajectory by selecting the required type of content. Dynamic content formation in the process of studying the course not only reduces the risks associated with reducing the effectiveness of information perception, but also allows the learner to motivate the study of the material. The proposed algorithmic solution allows not only to choose the type of content in terms of the form of its perception, but also to determine the necessary block (theory or practice).

To conduct the formation of the educational trajectory as a neural structure, the Kohonen network was chosen, since it most effectively performs clustering and classification of objects. An equally important factor is the visualization of the results, which allows, at an early stage, to improve understanding of the structure and nature of the data, and further refine the neural network model. Due to the peculiarities of virtual network functions, support for classification in the Kohonen network can be used to identify homogeneous elements, which will further optimize their choice for filling the course. The training of the Kohonen network is carried out by the method of successive approximations. Starting with a randomly chosen initial location of the centers, the algorithm gradually improves it so that it captures the clustering of training data. Another advantage of the Kohonen network is the ability to identify new clusters. The trained network recognizes clusters in training data and assigns all data to one or another cluster. If the network then meets a data set that is unlike any of the known samples, it will independently identify a new cluster of elements. This feature is very relevant, since it allows you to enter into the online course new types of objects without actually changing the algorithms for their assignment to students.

The principle of constructing a neural network system for optimizing the choice of course elements is as follows. Based on the data received from the training systems, we developed a number of criteria that can not only identify the course element, but also determine the need for its replacement. The criteria are formulated so that the answer can always be presented in binary form, that is, 1 - "Yes" or 0 - "No". Based on the obtained data, a signal vector $E = \{e_1, e_2, ..., e_n\}$, which is fed to the input of the neural network. The neural network is a two-dimensional matrix of neurons of dimension n (the number of inputs of each neuron) per m (the number of neurons). The number of inputs of each neuron is determined in relation to the previously established number of criteria. The number of neurons m coincides with the required number of partition classes, which corresponds to the number of unique course elements available on the digital educational platform. The significance of each of the inputs to the neuron is characterized by a numerical value called the weight, and is given in the form of a matrix X where the elements of the given matrix are the vectors of the weighting coefficients of the bonds $x^{i,j} = \left\{x_1^{i,j}, x_2^{i,j}, \cdots, x_n^{i,j}\right\}$.

The Kohonen network consists of three layers of neurons. The basis of the network is the hidden layer of Kohonen. However, in order to obtain results for the simultaneous identification and assignment of the course element in the study, we proposed a modified scheme of the output neurons of the Kohonen network (Fig. 1).

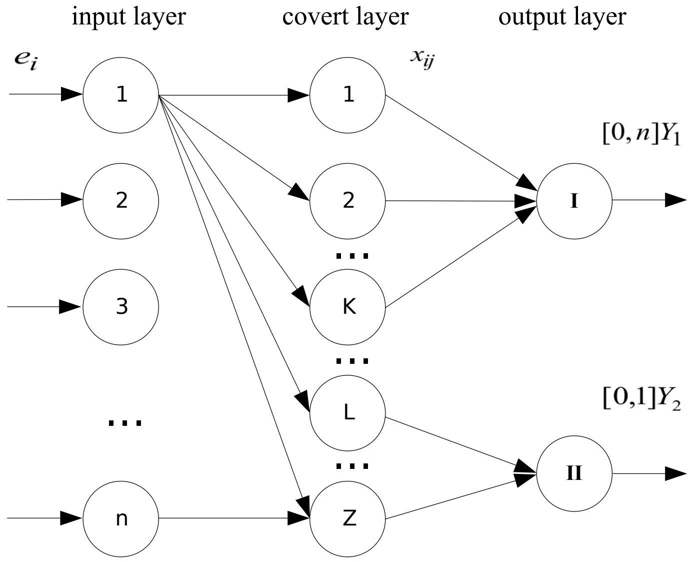

Figure 1: The neural network model

The hidden layer of the Kohonen neural network is proposed to be divided into two sets. The first set of neurons $[1...K]$ is responsible for identifying the element of the course assigned to the learner in the course of the course. During the operation of the neural network, by changing the input weights on the output layer, it activates the linear function Y_1, which takes the values $[0, n]$. At the same time, 0 means that the studied element of the course refers to the content of a non-critical value in the course of studying a particular course. For example, technical information or guidelines for working with the most digital educational platform. The numbers from 1 to n correspond to the specific course element identified by the neural network model. The second set of neurons $[L...Z]$ analyzes the load state of the course element being examined and at the output initializes the function Y_2, taking values $[0, 1]$, where 0 element does not require replacement for the trainee, -1 - means that the content needs to be changed.

To train the neural network in the framework of the study, the method of multi-page training was used. From the mathematical point of view, the training of neural networks is a multiparameter problem of nonlinear optimization. In the classical method of back propagation of an error (single-mode), the training of neural networks is considered as a set of one-criterion optimization problems. The criterion for each task is the quality of the solution of one example from the training sample. At each iteration of the backward propagation algorithm, the parameters of the neural networks are modified to improve the solution of one example. Thus, in the learning process, one-criteria optimization problems are cyclically solved.

From the theory of optimization, it follows that in the solution of multicriteria problems, parameter modifications should be performed using several criteria at once. Moreover, one example cannot be confined to evaluating the changes in the values of parameters. In order to take into account several criteria, parameters are aggregated or integrated, which may be, for example, the sum, weighted sum or the square root of the sum of the squares of the solution estimates of individual examples.

In particular, in the present studies, the change in weights was carried out after checking the entire training sample, while the error function was calculated as:

$$E(w) = \frac{1}{2} \sum_{k=1}^{n_1+n_2} \left(d^k - y^k\right)^2 , \quad (4)$$

where, k - number of the training pair in the training sample, $k = 1, 2, \cdots, n1 + n2$; n1 - number of vectors of the first class; n2 - number of vectors of the second class. As test tests show, training with the use of batch mode, as a rule, converges faster than learning by individual examples. The received information from the neural network is used to optimize the educational trajectory within the online course. To do this, a map of the optimal location of the course elements for a particular learner is formed, as well as possible changes, taking into account the individual characteristics of the learner and the information accumulated in the learning process. By analyzing the two maps and the heuristic forecasting algorithm, the learning management system makes a decision about adjusting the course structure and restructuring the training routes for the learner. At the same time, both maps are dynamic objects, formed not only as certain events occur in the digital educational platform, but also with a specified time interval, selected individually for each student.

4 Conclusion And Future Works

In the framework of the our research we outlined the solution to the complex task of forming an individual educational trajectory in MOOC based on the selection of the type of content. A mathematical tools based on the application of a genetic algorithm is described. We also describe in the algorithm for the intelligent control of individual educational trajectories in MOOCs based on the neural network approach. Now we are testing the portal (http://56bit.ru/) for online courses in computer science. As part of our research, it is established that the proposed algorithm increases the effectiveness and quality of the knowledge obtained in the course of mastering the online course chosen by the student's. This result are expressed in the achievement of student's planned learning outcomes, growth in the motivation of course participants, their satisfaction with the learning process.

Our further developments will be related to ensuring a permanent correction of the calculated individual educational trajectories based on student's academic history, achievements and rating. In the future, we apply clustering for the differentiation of students into homogeneous groups. This is necessary for the issuance of personal recommendations and selection of students for joint implementation of projects within the online course. To identify students who may not be able to cope with the performance of the assessment task, we will use forecasting methods.

5 Acknowledgements

The reported study was funded by RFBR according to the research project 18-37-00400 and 19-47-560011.

References

[Hon05] C. M. Hong, C. M. Chen, M. H. Chang Personalized Learning Path Generation Approach for Web-based Learning. *4th WSEAS Int. Conf. on E-ACTIVITIES*, 62–68, 2005.

[Hua07] M. J. Huang, H. S. Huang Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. *Procedia - Expert Systems with Applications*, 33(3):551–564, 2007.

[Bha10] A. M. F. Bhaskar, M. M. Das, T. Chithralekha, S. Sivasatya Genetic Algorithm Based Adaptive Learning Scheme Generation For Context Aware E-Learning. *Procedia - International Journal on Computer Science and Engineering*, 2(4):1271–1279, 2010.

[Han18] S. Azough, M. Bellafkih, El H. Bouyakhf Adaptive E-learning using Genetic Algorithms. *Procedia - IJCSNS International Journal of Computer Science and Network Security*, 10(7):237–244, 2010.

[Mar11] L. de-Marcos, et.c. Genetic algorithms for courseware engineering. *Procedia - International Journal of Innovative Computing, Information and Control*, 7(7):1–27, 2011.

[Yao14] X. Yao, D. Gong Genetic Algorithm-Based Test Data Generation for Multiple Paths via Individual Sharing. *Procedia - Computational Intelligence and Neuroscience*, 2014.

[Zha18] X. Zhang, L. Cao, Y. Yin Individualized Learning through MOOC: Online Automatic Test System Based on Genetic Algorithm. *16 Proceedings of the International Conference on Intelligent Information Processing*, 2018.

[Shm15] V. Shmelev, M. Karpova, A. Dukhanov An Approach of Learning Path Sequencing based on Revised Bloom's Taxonomy and Domain Ontologies with the use of Genetic Algorithms. *Procedia - Computer Science*, 66:711–719, 2015.

[Dwi05] P. Dwivedi, V. Kant, K. K. Bharadwaj Learning path recommendation based on modified variable length genetic algorithm. *Procedia - Education and Information Technologies*, 23(2):819–836, 2005.

[Gol89] D. E. Goldberg Genetic Algorithms in Search, Optimization, and Machine Learning. *Addison-Wesley*, 432, 1989.

[Bha10] M. Bhaskar, M. M. Das, T. Chithralekha, S. Sivasatya. Genetic Algorithm Based Adaptive Learning Scheme Generation For Context Aware E-Learning. *Procedia - International Journal on Computer Science and Engineering*, 2(4):1271–1279, 2010.

[Zap17] V. Zaporozhko, D. Parfenov, I. Parfenov. Approaches to the description of model massive open online course based on the cloud platform in the educational environment of the university. *International Conference on Smart Education and Smart e-Learning*, 75:177–187, 2017.

[Hua07] M. J. Huang, H. S. Das, M. Y. Chen. Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. *Procedia - Expert Systems with Applications*, 33(3):551–564, 2007.

[Par18] D. Parfenov, V. Zaporozhko. Developing SMART educational cloud environment on the basis of adaptive massive open online courses. *Conference Internationalization of Education in Applied Mathematics and Informatics for HighTech Applications*, 2093:35–41, 2018.

[Mar11] L. de-Marcos, J. J. Martinez, J. A. Gutierrez, R. Barchino, J. R. Hilera, S. Oton, J. M. Gutierrez. Genetic algorithms for courseware engineering. *Procedia - International Journal of Innovative Computing, Information and Control*, 7(7):1–27, 2011.

[Azo10] S. Azough, M. Bellafkih, El H. Bouyakhf. Adaptive E-learning using Genetic Algorithms. *Procedia - IJCSNS International Journal of Computer Science and Network Security*, 10(7):237–244, 2010.