

A genetic-algorithm approach for forming individual educational trajectories for listeners of online courses

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

Irina P. Bolodurina
Department of Applied Mathematics
Orenburg State University
Orenburg 460018, Russia
prmat@mail.osu.ru

Denis I. Parfenov
Faculty of Distance Learning Technologies
Orenburg State University
Orenburg 460018, Russia
parfenovdi@mail.ru

Abstract

One of the main directions for further improvement of the online courses is to provide complex personalization. The need for personalization of learning is a reflection of the natural for mankind desire for an individual approach to personal needs, preferences, and opportunities. A serious disadvantage of the online courses is the lack of an individual and differentiated approach to each student due to a pre-determined learning route in typical courses. In the present work, a genetic algorithm is proposed that allows you to form an optimal learning route, designed to meet the personal educational needs and individual capabilities of each listener of the massive open online courses. The results of a computational experiment and examples of individual trajectories formed on the basis of the proposed algorithm are presented.

1 Introduction

The individual educational trajectory in massive open online courses (MOOC) is a realization way of individual educational needs and abilities of students, their right to choose their personal development and self-improvement path [Sun15]. We define an individual educational trajectory as a personal path to realize the personal potential of each MOOC listener [Par18, Zap17]. There are several ways to realize an individual educational trajectory. For example, through the use of various educational technologies (tertiary differential education, problem-based learning, game-based learning, portfolio and others) or personalization technologies in MOOC (inquiry-based learning, personal recommender system, and others) [You15, Han18]. Another way is to form an individual learning route, which is a sequence of elements of the training activity of a particular student at some fixed stage of the study on the online course.

Copyright © by the paper's authors. Copying permitted for private and academic purposes.

In: Marco Schaerf, Massimo Mecella, Drozdova Viktoria Igorevna, Kalmykov Igor Anatolievich (eds.): Proceedings of REMS 2018 – Russian Federation & Europe Multidisciplinary Symposium on Computer Science and ICT, Stavropol – Dombay, Russia, 15–20 October 2018, published at <http://ceur-ws.org>

Purposefully designed an individual learning program is a technological tool for the implementation of an individual learning route. Individual learning routes for MOOC listeners differ not only in terms of volume but also in the variability of the forms of presentation of the electronic learning content. This is due to the individual learning styles of students and, accordingly, their activities used in the study of the same learning object. In our opinion, it is impossible to design an individual learning route in advance, as it must reflect the dynamics of learning, revealing it in movement and change. Such an approach will allow timely making necessary adjustments to the educational process implemented on the basis of MOOC. For example, to fill certain gaps in the knowledge and skills of the course listeners, or vice versa, to speed up the learning process or deepen the learning program.

The task of our research is to construct an optimal individual educational trajectory based on a genetic algorithm that is as close to the real possibilities and features of each listener of the course as well as corrected, if necessary, in real time. The remainder of this paper is organized as follows. In section 2, we present the results of a literature review devoted to the consideration of various approaches to the formation of an individual educational trajectory based on genetic algorithms. In section 3, we disclose the problem of the formation and implementation of an individual educational trajectory based on genetic algorithms. A mathematical model of the form of the optimal educational trajectory in the massive open online courses. Section 4 deals with the description of the practical implementation of the proposed model and the evaluation of the results obtained.

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 [Hon05] 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.

A group of researchers from Pondicherry University proposed to generate an adaptive learning scheme. The proposed approach allows to take into account the context-dependent content of learning. Depending on the educational goals and intentions of the learner, the most appropriate content is selected, which can be represented by three different types: Media, Presentation, Content. To select a particular type of content, researchers suggested using a genetic algorithm. On the basis of the data obtained, a learning path is drawn up, which best corresponds to the learner's intentions [Bha10].

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 curriculums 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 their 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 [Azo10].

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].

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 Problem formulation and implementation

As part of our study, MOOC has a modular structure consisting of a certain number of units. Within each unit, there are learning objects (LOs) of different types (Table 1), which are the structural components of the course electronic learning content [Zap17]. A certain set of LOs provides the formation of one or more relevant competencies.

It is known that each learner of the course has its own learning style [Zap06]. Researchers distinguish the following 4 types of students, differing in the dominant style of learning: Visual learners ("V"), Aural learners ("A"), Read-write learners ("R"), Kinesthetic learners ("K"). To what type each of the MOOC listeners belongs, we identify at the beginning of the learning process, using the VARK methodology [Fle95]. So, in our work, each listener of the course (as an object under study) is characterized by the following input parameters (a set of attributes characterizing the state of the given object), which are presented in Table 2.

We distinguish four generalized groups of content types depending on the dominant learning style (Table 1). For example, the first group consists of the types of content most suitable for students with the dominant modality "Visual". It is established that students can also have mixed modalities. Therefore, we propose to form a course with different types of content, but at the same time taking into account the revealed dominant modality as much as possible.

Table 1: Composition of four generalized groups of different types of content

Type of electronic learning content	Dominating Learning Styles	The designation of the content type	The attribute value coefficient	Relative attribute weight (b)
Group 1. Types of content most suitable for students with the dominant modality "Visual" (G1)				
Presentations (slides)	V	LO ₁	μ ₁	1
Infographics, illustrations	V	LO ₂		
Webinars	V	LO ₃		
Video lessons	V	LO ₄		
Group 2. Types of content most suitable for students with the dominant modality "Aural/Auditory" (G2)				
Audio conferencing	A	LO ₅	μ ₂	1
Audio notes	A	LO ₆		
Audio lessons	A	LO ₇		
Workbooks audio	A	LO ₈		
Group 3. Types of content most suitable for students with the dominant modality "Read/write" (G3)				
Glossaries	R	LO ₉	μ ₃	2
Reading	R	LO ₁₀		
Quizzes (or tests)	R	LO ₁₁		
Assignments	R	LO ₁₂		
Group 4. Types of content most suitable for students with the dominant modality "Kinesthetic" (G4)				
Video and educational games	K	LO ₁₃	μ ₄	2
Virtual laboratories	K	LO ₁₄		
Interactive learning	K	LO ₁₅		
Workshops	K	LO ₁₆		

Thus, a number of LO from the list of each group must be present in each unit. Accordingly, for each listener, a unit must be dynamically formed, consisting of LO, mainly corresponding to its learning style.

To establish a representative correlation of different types of content (learning objects) of a particular unit, depending on learning style, 15,457 respondents were surveyed. The use of the VARK methodology allowed an analysis of the real situation.

Based on the results of the survey, we will determine the ratio of different types of content in a specific online course for each type of student (Table 3). Then the sum of the content types ratio of the different groups for

Table 2: Characteristics and values of attributes for a set of students

Attribute Name	Attribute (parameter)	Possible Attribute Values	The attribute value coefficient	Relative attribute weight (ν)
a_1	Gender	Female	$a_{1,1}$	1
		Male	$a_{1,2}$	
a_2	Age group	under 18	$a_{2,1}$	2
		19-25	$a_{2,2}$	
		26-34	$a_{2,3}$	
		35-44	$a_{2,4}$	
		45-54	$a_{2,5}$	
		55+	$a_{2,6}$	
a_3	Learning Style	Visual learners	$a_{3,1}$	3
		Aural learners	$a_{3,2}$	
		Read-write learners	$a_{3,3}$	
		Kinesthetic learners	$a_{3,4}$	

each type of learner should be equal to one $\mu_1 + \mu_2 + \mu_3 + \mu_4 = 1$. Varying the ratio of $\mu_1, \mu_2, \mu_3, \mu_4$ in the overall content structure gives different sets of LOs in the individual learning route.

Table 3: The ratio of different types of content in each unit

Types of students (by VARK)	The weight of each type of content in the course structure			
	G_1	G_2	G_3	G_4
	μ_1	μ_2	μ_3	μ_4
Visual learners	0.31	0.21	0.26	0.22
Aural learners	0.25	0.31	0.22	0.22
Read-write learners	0.24	0.18	0.34	0.24
Kinesthetic learners	0.22	0.23	0.24	0.31

Completion of the study of each unit is accompanied by the performance of a summative test, the results of which allow one to draw a conclusion about the success of learning process or the prevalence of difficulties in the course student.

Thus, the individual learning route in MOOC is a varied set of learning objects of different types for each of the units. Their list is formed and adjusted in real time mode when the listener moves from stage to stage (from one unit to the other).

3.1 Mathematical model of the problem

We created a model for the formation of the individual educational trajectory in the online course. Let us presented the initial data for solving the claimed problem with the help of the mathematical tools of the genetic algorithm [McC05].

Having analyzed the subject area of the task, we have identified the following tuple, characterizing the formation process of the individual educational trajectory (IET).

$$IET = (S, C, P), \quad (1)$$

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

Each unit of the MOOC contains a specific set of content groups. Then let $G = g_1, \dots, g_n$ 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 1). $Unit_x = G_1, \dots, G_4$.

Then $P = P_1, \dots, P_n$ 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 (according to the Table 1). 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 is presented in Tables 1 and 2, respectively. 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 attributes values in order of increasing importance.

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

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

where μ_{h,s_k} - attribute coefficient value μ_h for unit x depending on the particular type of student s_k, b_h relative weight of attribute g_h for unit (Table 1).

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:

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

where a_{y,s_k} - attribute coefficient value a y for student s_k, ν_h - relative weight of attribute a_y (Table 2)

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 find an acceptable (optimal) individual learning route P in a cloud-based learning environment (depending on parameters $(a_1, \dots, a_3, \nu_1, \dots, \nu_4)$), we use the genetic algorithm.

3.2 Individual educational trajectory generation based on genetic algorithm

We consider a genetic algorithm that works with a population (a finite set of individuals). The set of optimized parameters is represented in the form of genes that form a chromosomal filament. In the chromosome of each individual, a possible solution of the problem is encoded. This algorithm consists of the following steps:

Step 1. Initialization (formation) of the initial population from P chromosomes. The population is a collection of several vectors P. The size of the population is set before the genetic algorithm begins work. The individual is one element of the vector P. The gene is an element of $LO_{i,j}$ from the vector P. In our model, the chromosome consists of LO genes, in which the alleles of each of the genes are the values of $\{0, 1\}$.

Step 2. Calculate the fitness function of the chromosome in the population F(P).

The objective function numerically characterizes the result of selecting an individual educational trajectory in MOOC by the following formula:

$$F(P) = F(P)_{max} - \sum_1^x (W_{Unit_x} \cdot W_{s_k} \cdot T_x(S_k) \cdot Z(P_x)), \quad (4)$$

where P vector of selection of individual learning route; W_{Unit_x} the weight of each unit in the course; W_{s_k} the weight of each student; $T_x(S_k)$ student test score in each unit; $F(P)_{max}$ maximum value of the objective function; $Z(P_x)$ function of formation a set of LOs.

Step 3. Selection of the best individuals from the current population (two parent chromosomes) for further crossbreeding using one of the selection methods. Selection: the fittest individuals have the best chance of reproducing.

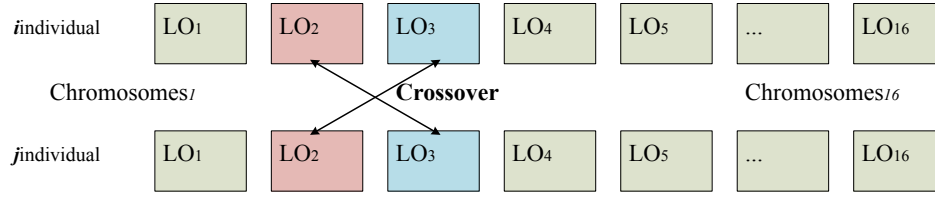


Figure 1: Crossover operator in a genetic algorithm.

Step 4. The use of the genetic operator crossover. Crossover: exchange genetic material between two individuals (see Fig. 1). Creation of a new population of descendants on the basis of the original one using a crossover.

Step 5. The use of the genetic operator mutation. Mutation: randomly change part of the genetic material (see Fig. 2). Creation of a new population of descendants on the basis of the original with the help of a mutation of individuals (descendants) with a certain probability.

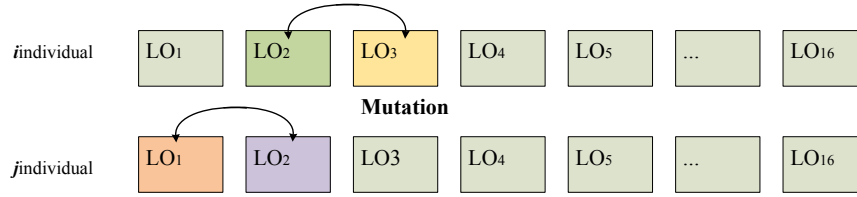


Figure 2: Mutation operator in a genetic algorithm.

Step 6. Repeat steps 3-5 until a new generation of the population containing n chromosomes is generated.

Step 7. Repeat steps 2-6 until the end-of-process criterion is reached - the "best" chromosome (the optimal solution of the problem is found).

The criteria for termination of the genetic algorithm are as follows: obtaining a solution of the required quality; the solution falls into a deep local optimum of the objective function; search time expired.

4 Experimental results

In this section the results of a simulation study are presented. Using the built-in functions of MATLAB, we implemented a genetic algorithm with the following experimental parameters. The size of the population, we have established 50 individuals. Each chromosome is represented as a binary code. The probability of a mutation is 0.05. The probability of crossing-over is 0.8. The Table 4 illustrates individual learning routes, which are obtained from the results of the experiment. The experimental realization of our algorithm was carried out for Information Technology MOOCs for technical specialties at the university.

The LO value is "0" if this learning object is not included in the individual learning route. The LO value is "1" if this learning object is present in the individual learning route.

5 Conclusion

In this article, we introduced a new algorithm that allows forming individual educational trajectories of MOOC listeners. This algorithm is proposed for the cloud educational platform, which implements the concept of personalized learning. The mathematical tools of the genetic algorithm are used in this proposed solution. The created algorithm is able to find the optimal set of course learning objects that constitute an individual learning route. The results of the computational experiment show that the proposed algorithm is able to find solutions that are very close to optimal solutions and in most cases are identical to them.

6 Acknowledgements

The research was conducted with the support of the Russian Foundation for Basic Research (project no. 18-37-00400).

Table 4: Examples of individual learning routes for one unit formed on the basis of the genetic algorithm

Listener of MOOC	Gender	Age group	Dominating learning style	Individual learning route
Student 1	Female	19-25	Aural	0 0 1 1 1 0 1 1 1 0 0 1 0 0 0 1 $LO_3LO_4LO_5LO_7LO_8LO_9LO_{12}LO_{16}$
Student 2	Male	19-25	Aural	0 1 0 1 1 1 0 1 0 0 1 1 1 0 0 0 $LO_2LO_4LO_5LO_6LO_8LO_{11}LO_{12}LO_{13}$
Student 3	Female	19-25	Read-write	0 0 1 0 0 0 0 1 1 0 0 1 1 0 1 0 $LO_3LO_8LO_9LO_{12}LO_{13}LO_{15}$
Student 4	Female	19-25	Kinesthetic	0 0 0 1 1 0 0 0 1 0 0 1 1 1 1 0 $LO_4LO_5LO_9LO_{12}LO_{13}LO_{14}LO_{15}$
Student 5	Female	19-25	Visual	1 1 1 0 0 1 0 1 1 1 0 0 1 0 0 0 $LO_1LO_2LO_3LO_6LO_8LO_9LO_{10}LO_{13}$
Student 6	Male	19-25	Kinesthetic	0 0 1 0 0 0 1 1 0 1 1 0 1 1 0 0 $LO_3LO_7LO_8LO_{10}LO_{11}LO_{13}LO_{14}$
Student 7	Female	19-25	Read-write	0 0 0 1 0 1 0 0 0 1 1 0 1 0 1 0 $LO_4LO_6LO_{10}LO_{11}LO_{13}LO_{15}$
Student 8	Male	19-25	Visual	0 1 0 1 0 0 1 0 0 1 1 0 0 1 0 0 $LO_2LO_4LO_7LO_{10}LO_{11}LO_{14}$

References

- [Sun15] A. S. Sunar, N. A. Abdullah, S. White, H. C. Davis. Personalisation of MOOCs: The State of the Art. *7th International Conference on Computer Supported Education*, 1:88–97, 2015.
- [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.
- [You15] A. M. F. Yousef, A. S. Sunar. Opportunities and challenges in Personalized MOOC Experience. *ACM WEB Science Conference 2015, Web Science Education Workshop 2015*
- [Han18] H. Yu. Han, C. Miao, C. Leung, T. J. White. Towards AI-powered personalization in MOOC learning. *Science of Learning*, 15:1–5, 2017.
- [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.
- [Zap06] A. Zapalska, D. Brozik. Learning styles and online education. *Campus-Wide Information Systems*, 23(5):325–335, 2006.
- [Fle95] N. D. Fleming. I’m different; not dumb. Modes of presentation (VARK) in the tertiary classroom. *1995 Annual Conference of the Higher Education and Research Development Society of Australasia, Research and Development in Higher Education*, 18:308–313, 1995.
- [McC05] J. McCall. Genetic algorithms for modelling and optimization. *Computational and Applied Mathematics*, 184:205–222, 2005.
- [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.
- [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.

- [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.
- [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.
- [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.