Development and Research of Algorithms for the Formation of an Optimal Individual Educational Path for Online Courses

Parfenov D.I.[*] Zabrodina L. Zaporozhko V.

Orenburg State University, Orenburg, 460018, Russia
[*]Corresponding author. Email: parfenovdi@mail.ru

ABSTRACT

An important condition for making an effective decision in the field of e-learning is the analysis of data from participants in the educational process at its various stages. Currently, the volume of data circulating in a digital educational environment that supports working with online courses is growing exponentially. This is facilitated by the rapidly growing demand for open education. In such an environment, for the effective study of materials, it is necessary to carry out a personified process control of the flow of knowledge. One way is the formation of individual educational paths. In the framework of this study, a model for the formation of the individual educational path of the student based on the solution of the problem of minimal coverage of the graph is proposed. As the basic parameters of the model, a competency-based approach is applied, superimposed on the structural description of the relationships of the individual components of the online course. In the framework of the study, genetic and greedy algorithms were used to form an educational path, as well as a comparative analysis of the approaches presented.

Keywords: genetic algorithm, greedy algorithm, discrete optimization, individual educational path, coverage problem

1. INTRODUCTION

The question of the emergence of individual educational routes and their application in the practice of modern education is associated with the introduction of the principle of an individual approach to the development of students' abilities and knowledge. The main idea, in this case, is aimed at the awareness of the students of their abilities and the formation on their basis of a unique learning path of various fields of knowledge [1,2].

In the framework of this study, the problem of constructing the most optimal educational path in terms of the quality of knowledge gained and the time spent on learning is considered. Under the quality of knowledge obtained in the learning process, this study examines the set of competencies acquired by students.

The task of optimizing an individual educational path can be reformulated into the problem of least coverage as follows: within a specific discipline with competencies tied to it, it is required to find the minimum number of training modules in such a way as to cover the entire list of necessary competencies and build an individual educational path for them, in mastering which the student will receive a sufficient amount of knowledge, skills, abilities, including the mandatory learning outcomes.

1.1. Related Work

Research and development of algorithms for solving the generalized problem of minimum coverage are engaged around the world.

The team of authors from the University of Bologna in their study considers issues of solving the problem of covering sets with metaheuristic approaches: the impossibility of solution and redundancy of sets [3]. In this paper, the problem of covering sets is reformulated into the problem of unconditional optimization, which eliminates the need to solve impossibility problems and establish redundancy problems. Also, the existing greedy algorithm for the problem of covering sets has been adapted to the new formulation.

The authors of the study [4] considered the NP-complete problem of covering an undirected graph and proposed using a modified genetic algorithm to solve it. The effectiveness of the proposed method in the work is confirmed by a comparative analysis with other existing algorithms in a series of experiments.

In the framework of research [5], the genetic algorithm for solving the problem of finding the minimum coverage with an inhomogeneous cost is described in detail. The results of the study showed that the developed algorithm obtains the optimal solution for the problems of covering medium and large dimensions.

The publication [6] solved the problem of covering with interval weights. The authors of the study proposed a modification of the greedy algorithm for finding the joint approximate solution, understood as the approximation of the united optimal solution with guaranteed accuracy.

In the study [7], the authors use an approximate algorithm to find a solution to the vertex coverage problem that is close to the optimal one. The paper presents a greedy search algorithm for almost optimal coverage, which provides a search for a solution in polynomial time. The proposed approximation algorithm gives an optimal solution in most cases.

To construct educational path, several researchers use data on the individual characteristics of students. In the study [8], the authors propose two approaches to solving this problem, one of which is based on the use of methods based on neural networks.

In the study [9], the authors propose a new method for the formation of the educational path, based on assessing the effectiveness of the already studied disciplines using a dynamic control model. At the heart of the proposed model, the authors use a competency-based approach in combination with the use of modern information technologies for the analysis of learning outcomes.

In a study [10], the authors in practice propose to apply the concept of individualization of educational activity, taking into account the characteristics of various types of cognitive processes that occur in the process of mastering educational programs.

Zoja Raud et al. used the ontological design of educational path [11]. The method proposed by the authors allows us to determine the basic set of competencies for the chosen field of knowledge of the student. Based on the data obtained, the authors propose to form a consistent plan for studying the cycle of disciplines.

The main task of constructing an individual educational path for the student is to reduce the risks of not completing the selected courses, as well as forecasting the time of their development [12].

A review of the studies showed that at the moment there are many algorithms for constructing educational path, but some modifications are needed to apply them to particular optimization problems.

1.2. Our Contribution

In the framework of this study, a model for the formation of the individual educational path of the student based on the solution of the problem of minimal coverage of the graph is proposed. As the basic parameters of the model, a competency-based approach is applied, superimposed on the structural description of the relationships of the individual components of the online course. In the framework of the study, genetic and greedy algorithms were used to form an educational path, as well as a comparative analysis of the approaches presented.

1.3. Paper Structure

The article is organized as follows. Section 2 presents a formalization of the problem and briefly describes how to build and optimize an individual educational path by using a genetic algorithm and the greedy algorithm. Section 3 is devoted to experiments with real data from the history of studying a number of online courses. Section 4 concludes the paper and presents direction for future research.

2. THE MODEL FOR BUILDING AND OPTIMIZING AN INDIVIDUAL EDUCATIONAL PATH

2.1. Problem statement

The educational path of any discipline/field of knowledge can be represented in the form of training modules, each of which represents a cyclical implementation of educational activities, namely: the study of theoretical material (lectures) and the implementation of practical tasks (Figure 1). The number of training modules should be optimal, consistent with the number of topics of the studied discipline, as well as with a minimum set of competencies, which should be formed at the end of the study of this discipline.
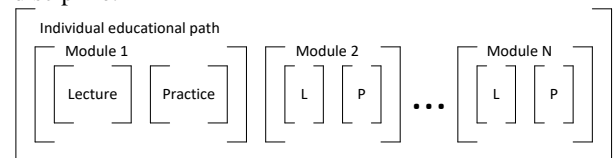

Figure 1 The individual educational path in the context of the course

Each training module has a set of competencies necessary for passing this particular training module and a set of competencies that will be formed after passing the study of the material (Figure 2). Also, the time spent on studying a particular training module is fixed (average value).
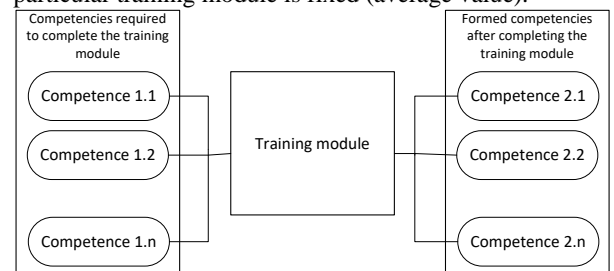

Figure 2 The procedure for the formation of competencies based on the results of the course

Thus, it is necessary to build an individual educational path so that the competencies necessary for passing each module are formed at the previous stages, or their presence was assumed before studying the chosen discipline (Figure

3). In this case, it is necessary to take into account the initial training of the student (existing competencies) and the time spent on passing the selected educational path.

Where Kx.y.z denotes a separate competency, x is a number, takes the values 1 or 2, where 1 is a set of competencies necessary before completing the training module, and 2 is a set of competencies that will be formed after passing the training module;

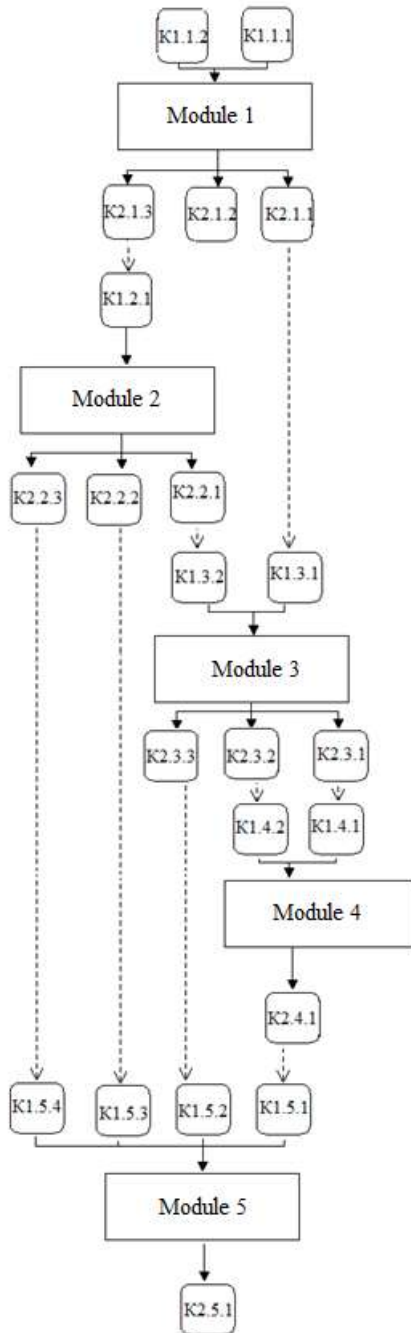y is the number of the training module;

z is the competency number.

Figure 3 The model for competency formation

So, within the framework of a particular discipline with competencies tied to it, it is required to find the minimum number of training modules in such a way as to cover the entire list of necessary competencies and build an individual educational path along which, when mastered, the student will receive a sufficient amount of knowledge, skills, and abilities, including the required result learning.

This problem relates to the problems of the minimum coverage and the minimum vertex coverage. They have wide applied value in the theory of complex construction, in the development of their software and software. This task and related problems have various applications in adaptive query processing and resource allocation. The main requirement for the algorithms for solving these problems is the high efficiency of the solution and ensuring the minimum possible error of these solutions.

The initial data of the problem of covering a set is:

The semantic network is a strongly connected acyclic directed graph (GD).

Parameters:

The field of knowledge of an individual course is D;

Course Semantic Network: $GD = \{VD, ED\}$, where

VD - set of vertexes (competencies);

ED - set of graph links.

Before building an individual educational path, it is necessary to determine the initial level of knowledge of students. For this, it is planned to conduct an input test, based on which a student's semantic network of knowledge will be built.

The student's semantic knowledge network is a strongly connected acyclic oriented graph (GL).

Parameters:

Student L;

Learner's semantic knowledge network $GL = \{VL, EL\}$, where

VL – set of vertexes (competencies);

EL – set of graph connections;

$VL \subseteq VD, EL \subseteq ED$;

The graph GL is a connected subgraph of the graph GD.

Suppose that a graph describing all the links in the course and their interconnections can be reduced to a certain minimum graph in connection with the student's competencies. The task of forming the minimum number of vertices in the graph of the course under study can be solved using the general scheme of the genetic algorithm.

2.2. The genetic algorithm for optimizing an individual educational path

We define a population of size N as a set of different individuals, each of which is a family of coatings. Thus, each set represents the vertices of the graph of the studied course, satisfying the admissibility condition. By admissibility, we mean such a truncated set of vertices that ensures the formation of all necessary competencies. Covers of the initial population are built randomly. Each set is a sequence of values "0" and "1" for each training

module, where "0" - do not include the module in the route, "1" - turn on the module in the route.

Optimal is a set of vertices that contains the smallest possible number of vertices. In case of coincidence of the number of vertices for different sets of vertices (different coatings), the optimal one is considered to have less than the total time required to complete the training blocks.

The fitness function for each cover/set of vertices is the ratio of the number of vertices/training modules used to the total number of training modules.

We carry out the mutation operation as removing a random vertex from the general coverage. If a mutated set of vertices falls out of the admissibility region, a rollback to the original set occurs.

The crossing of individuals of the population - different sets of coatings are defined as two-point, i.e. 2 random points are generated (from 2 to N - the number of training modules). A new set of vertices is obtained by replacing the information about the choice of a particular vertex (training module) in 2 individuals with the values of the vertex selection of 1 individual (only the module selection values for vertices whose numbers fall in the interval of the generated values are replaced).

Selection (selection of the best individuals of the population) is carried out saving a certain percentage of the best individuals (sets of training modules of the course, providing coverage with the highest learning speed) in the new generation.

The optimization process will go on until the best set of vertices changes over several generations of gen (Figure 4).
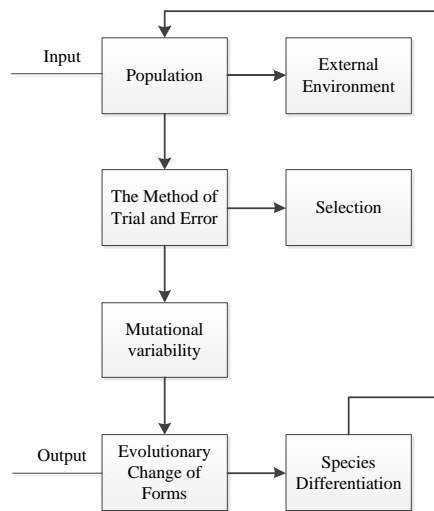

Figure 4 The genetic algorithm for optimizing an individual educational path

Thus, in order to determine the most optimal overlap of the graph of the studied course, it is possible to optimize the genetic algorithm presented in the scheme (Figure 4) based on the functions described above.

2.3. The greedy algorithm for building an individual educational path

Consider a greedy algorithm for solving the set covering problem. The solution constructed using this algorithm converges more slowly than the solution of the integer programming problem in a logarithmic number of times. Consequently, with an increase in the dimension of the problem, the accuracy of the solution decreases, but rather slowly. In this regard, the proposed approach can be considered useful.

The greedy algorithm works as follows: at each iteration, a set is selected that provides maximum coverage for elements that are not overlapped in the previous steps.

As part of the solution to the problem of constructing and optimizing an individual educational path, it is necessary to adjust the classical greedy algorithm.

Firstly, it is necessary to take into account the initial training of the student (existing competencies). In this regard, the initial set of overlapping competencies (formed in the process of studying the educational course) may not be empty but contain many competencies.

Secondly, it is necessary to take into account the time taken to complete the selected educational path. To make a choice and add the least time-consuming training module to the individual educational path, we introduce an additional check: if the greatest coverage of competencies in one iteration provides several training modules, then we add a less time-consuming to the educational path.

Greedy approximate algorithm:

Input data:

U — given set of all elements;

F — family of subsets U;

U_0 — Initial set of learners' competencies mastered before starting the course;

Step 1. $X \leftarrow U$

Step 2. $C \leftarrow U_0$

Step 3. while $X \neq \varnothing$ do

    1. select $S \in F$ with the greatest $|X \cap S|$ - the number of new competencies covered by the training module S;

    2. if there are several training modules S with the same maximum value $|X \cap S|$, then select the module with the least learning time Δt;

    3. $X \leftarrow X \setminus S$;

    4. $C \leftarrow C \cup \{S\}$;

Step 4. return C

3. EXPERIMENTAL RESULTS

As part of an experimental study, a comparative analysis of the genetic and greedy algorithms for constructing the optimal individual educational path for a number of online courses has been carried out. The developed algorithms were tested during the educational process for the

implementation of blended learning using online courses in four areas of training at the Orenburg State University in the following academic disciplines: "Modern Information Technologies", "Technologies for the Development of Computer Learning Tools", "Modern Means for Assessing Learning Results", "Use of modern information and communication technologies in education", "Simulation modeling", "Designed and management of intelligent systems", "Information Technologies".

For the experiment, we used information on 419 students in four areas of study that chose to study the above-listed online courses. The initial data for the analysis is the information obtained from the data warehouses of the online learning platform (Moodle). It was found that the developed algorithms and methods can reduce wasted time by 20-25% when developing an online course and increase the effectiveness of training, expressed in the growth of motivation and self-discipline of students, in their satisfaction with the process of acquiring the planned learning outcomes, as well as in increasing 32% of the number of students who successfully completed the course.
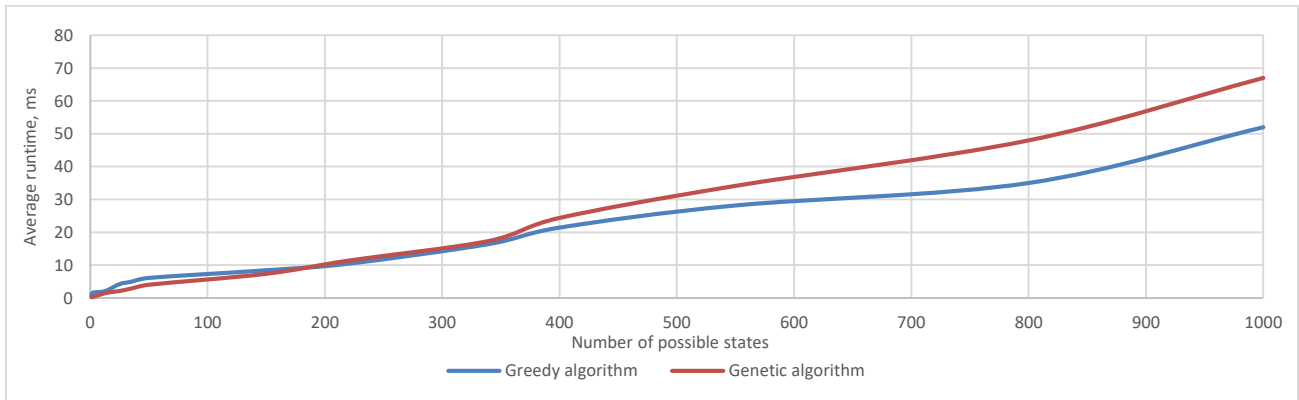

Figure 5 The dependence of the average operating time of the algorithms on the number of possible solution states.
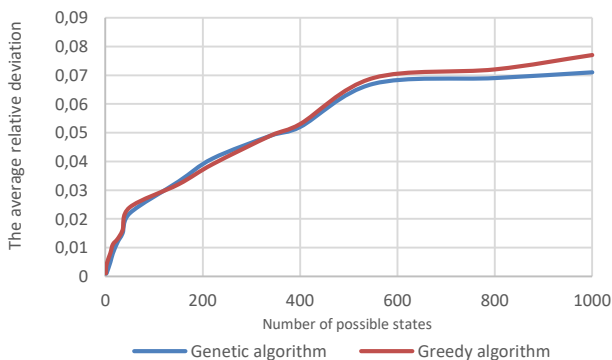

Figure 6 The dependence of the average relative deviation from the number of possible solution states

Figure 7 The dependence of the amount of false "lack of solution" of algorithms on the number of possible states of solution

For each online course, we compare the operating time of the genetic and greedy algorithms for constructing individual educational path (Figure 5). As expected, for training courses with construction options of more than 380 individual paths, the greedy algorithm finds an answer faster than the genetic algorithm on the same data (this is due to the simple enough structure of the algorithm). However, if there are fewer options for individual educational paths, the genetic algorithm converges faster.
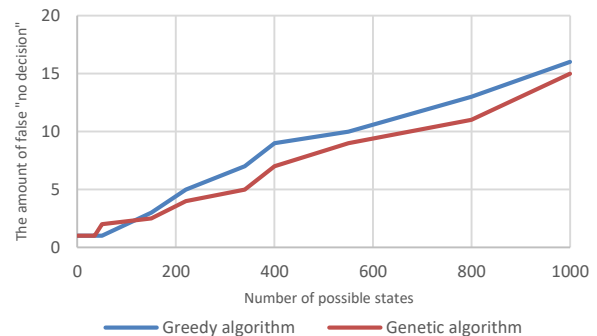
In addition, it is worth noting that with an increase in the number of solution states, the relative deviation of the result from the minimum and the number of false "lack of solutions" of the problem for each of the presented algorithms increase (Figures 6-7). However, the number of false "lack of solutions" genetic is much lower than the greedy algorithm. In this regard, for further use in the automatic generation of optimal educational path of training courses, a genetic algorithm will be used to improve the quality of individual courses being formed.

3. CONCLUSION

As a result of the study, genetic and greedy algorithms for the formation of the individual educational path of the student have been developed, which allow adaptive space-time planning of individual educational path of students of online courses. For training courses with construction options of more than 380 individual path, the greedy algorithm showed faster results, however, if there are fewer options, the genetic algorithm converges faster. In addition, the advantage of the genetic algorithm is the rather low values of the false response of the algorithm

The presented algorithms can be used by the recommender system to select the courses necessary for the formation of a given set of competencies, as they allow to optimize the study time (route) of individual blocks of disciplines mastered using a self-organizing educational environment. The developed algorithms allow differentiated for each student to quickly, efficiently and effectively form sets of socially demanded, general and professional competencies.

ACKNOWLEDGMENT

REFERENCES

[1] N.V. Khokhlova, L.D. Starikova Formation of Students' Individual Educational Trajectories by Means of the Elective Content. Yaroslavl pedagogical bulletin. vol. 4. (2018). pp. 73-79. DOI: https://doi.org/10.24411/1813-145X-2018-10106.

[2] E. Ivanova M., A.V. Vishnekov A computer design method of an effective educational trajectory in blended learning based on students' assessment. Education and Information Technologies, Springer, 2020, pp. 1–20. DOI: https://doi.org/10.1007/s10639-020-10109-3.

[3] A. Caprara, P. Toth, M. Fischetti Algorithms for the Set Covering Problem. Annals of Operations Research. vol. 98. (2000). pp. 353–371. DOI: https://doi.org/10.1023/A:1019225027893

[4] K. H. Han, C. S. Kim Applying Genetic Algorithm to the Minimum Vertex Cover Problem. The KIPS Transactions: PartB. vol. 15(6). (2008). pp. 609-612. DOI: https://doi.org/10.3745/KIPSTB.2008.15-B.6.609.

[5] M.K. Nguen Application of a genetic algorithm for the problem of finding coverage of a set. Proceedings of the institute for systems analysis Russian Academy of Sciences. vol. 33. (2008). pp. 206-219.

[6] A.V. Prolubnikov A greedy algorithm for solving the coverage problem with interval weights, in Problems of optimization and economic applications 2015. pp. 114-114.

[7] C. Dimri, K.C. Purohit, D. Pant, A greedy approach based algorithm for the vertex cover problem. International Journal of Scientific and Engineering Research vol. 4(3). (2013). pp. 1-6.

[8] V.V. Zaporozhko, D.I. Parfenov, M. Lapina, D. SoraDevelopment and Research of Algorithms for the Formation the Individual Educational Trajectories of Students in the Digital Educational Platform, in: Proceedings of SLET-2019 – International Scientic Conference Innovative Approaches to the Application of Digital Technologies in Education and Research, CEUR-WS, 2019. pp. 1-7.

[9] A. Mitsel, N. Cherniaeva Models, Methods, and Algorithms for Control over Learning Individual Trajectory. Handbook of Research on Estimation and Control Techniques in E-Learning Systems. IGI Global (2016). pp. 245-257. DOI: https://doi.org/10.4018/978-1-4666-9489-7.ch017.

[10] L. Konnova, L. Lipagina, G. Postovalova, A. Rylov, I. Stepanyan Designing adaptive online mathematics course based on individualization learning. Education Sciences, vol. 9(3). (2019) p. 182. DOI: https://doi.org/10.3390/educsci9030182

[11] Z. Raud, V. Vodovozov, E. Petlenkov, A. Serbin Ontology-based design of educational trajectories, in: 2018 IEEE 59th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), IEEE, 2018. pp. 1-4. DOI: https://doi.org/10.1109/RTUCON.2018.8659893

[12] F. Dalipi, A.S. Imran, Z. Kastrati MOOC dropout prediction using machine learning techniques: Review and research challenges, in: 2018 IEEE Global Engineering Education Conference (EDUCON), IEEE, 2018. pp. 1007-1014. DOI: https://doi.org/10.1109/EDUCON.2018.8363340