Automatic Generation and Evolution of Personalized Curriculum Based on Genetic Algorithm

https://doi.org/10.3991/ijet.v14i12.10812

Xiaocong Duan Guangdong University of Finance and Economics, Guangdong, China rmgyrrezptpv7@163.com

Abstract—In order to solve the problem of automatic generation and evolution of personalized curriculum, the method of using the genetic algorithm to realize the evolution of personalized learning content is proposed to satisfy the dynamic personalized needs of users. Through the research and implementation of personalized curriculum generation technology, first, the structures of curriculum generation, genetic algorithm and curriculum scene, as well as the method and technology of personalized curriculum generation and evolution system framework are described in detail. Then, the framework structure based on genetic algorithm is determined, and the user model is updated. Finally, experiments are carried out based on the genetic algorithm. The research on the experiment of automatic generation and evolution of personalized curriculum shows that the application of genetic algorithm in the process of curriculum generation and evolution makes students' learning content evolve with the change of their knowledge state in the process of learning, effectively promotes students' interest in learning, and improves learning efficiency and effect.

Keywords—Genetic algorithm, personalized learning, curriculum generation, curriculum evolution

1 Introduction

With the rapid development of modern communication technologies, multimedia technology, and Internet technology, network learners can obtain various learning resources in a variety of fast and convenient ways. However, some problems also appear, such as how to acquire the relevant knowledge that learners need from a large number of learning resources so as to complete learning a specific course from formal learning? How can learners acquire learning resources in accordance with their current learning status and learning objectives? This often cannot be accomplished by a learner's own ability. Instead, it needs professional teachers to make appropriate selection and compilation, and construct a specific network course in accordance with the individual needs and objectives of learners based on the correct assessment of learners' learning level, cognitive ability, and knowledge level. Therefore, personalized curriculum generation technology and method is a very important research direction in the field of e-learning.

The generation and realization of individualized learning resources can be regarded as an important step of individualized education, and the generation of individualized learning resources must be based on the evaluation of students' personality characteristics and learning process. However, learning is an individual cognitive process. Individualization is also a concept in psychology, sociology, pedagogy, philosophy, and other disciplines, which has considerable complexity. The key problem of online learning is how to extract, count, and analyze the key information and data in the process of students' learning in the current online learning environment, to get personalized characteristics of different students, and to generate personalized learning content accordingly. In the process of generating learning content, it is supposed to consider the teaching objectives, the requirements of different levels of the curriculum, as well as the teachers' syllabus and teaching content.

The research is to solve the problem of automatic generation and evolution of curriculums in distance network learning environment to meet the needs of personalized learning. When large-scale learners are in urgent needs of appropriate learning resources that can meet their individual needs, facing the lack of computer skills of many professional teachers and the needs for simple methods to solve the problem of network course construction, how to use algorithms in a large knowledge base and a large resource base to generate curriculum knowledge areas and suitable learning resources for formal learning? In order to adapt to the personalized needs in the learning process, how to update the course content dynamically? Based on these problems, a learning state tracking method based on personalized knowledge state graph is proposed. The genetic algorithm is applied to generate and evolve personalized curriculum content, which solves the dynamic updating problem of curriculum content.

2 Literature Review

The study of curriculum generation technology has always been one of the hot topics in the field of network learning based on the Web. Girgis et al. (2014) proposed a structure-oriented technology, which uses the genetic algorithm to automatically generate test paths for a set of courses covering all usage criteria [1]. For programs with loops, the proposed technique generates paths according to the subset criterion, which requires traversing zero, one and two loops paths. The set of paths generated by the proposed genetic algorithm can be passed to the test data generation tool to find program input that will execute these paths. Compared with the random test path generation technology, the effectiveness of the algorithm is verified by experiments. Yin et al. (2018) designed the research of curriculum generation based on an expert system and attempted to use the methodto separate the inference engine from a knowledge base. They also tried to solve the problem of curriculum generation using chromosome problems in graph theory, and the research work was deepened continuously [2]. Chen et al. (2014) proposed the evolutionary design based on the structural gene model of life cycle assessment and established the genetic model of structural life cycle assessment characteristics [3]. They also converted the variable length coding to equal length coding to achieve quantitative representation of structural information.

Moreover, they established the fitness function of structural life cycle assessment by using the analytic hierarchy process and designed genetic operators. The genetic model based on structural life cycle assessment realizes the evolutionary design of design schemes, reflects the life cycle evaluation characteristics of design schemes and improves the efficiency of the product design generation. Shi and Zhang (2014) analyzed and compared the advantages and disadvantages of genetic algorithm and evolutionary coding in solving system identification problems. They also proposed a new system identification method combining genetic algorithm and evolutionary coding. This method does not depend on the initial value of the population but still has strong stability. The simulation results showed the effectiveness and uniqueness of the method [4]. Yu et al. (2017), aimed at solving the slow evolution speed and premature convergence of the standard genetic algorithm, put forward an improved genetic algorithm based on multi-objective optimization model by improving gene coding, population initialization and genetic operators [5]. The improved genetic algorithm has better comprehensive execution speed than the standard genetic algorithm and improves the maximum fitness (0.79-0.84) and the average fitness (0.74-0.83). Based on the algorithm of time bitmap superposition matching, De et al. (2016) transformed the problem of course generation into the problem of course hour pattern searching based on time bitmap superposition matching, abstracted the course generation process, and eliminated the conflict of course generation by various backtracking methods [6]. Sahu et al. (2014), on the basis of priority automatic curriculum generation algorithm, used the idea of zero integration in operational research to transform these problems into physical four-dimensional space problems, which can reduce the width of space [7]. Xin et al. (2017) used artificial intelligence and decision support system to generate courses. These systems start from the school's teaching management and depend on the actual constraints, artificial-subjected and intelligence-assisted work [8]

To sum up, based on the above research, the methods and technologies used are introduced. Second, the structure of curriculum generation, genetic algorithm, relationship between curriculum generation and genetic algorithm, and personalized course generation and evolution system framework are discussed. Moreover, the results are analyzed in detail from the system framework structure and user model update. At last, the effects of mutation probability on average fitness and optimal fitness, and the influences of crossover probability on average fitness and optimal fitness are analyzed specifically. The advantage of the study is that the genetic algorithm is effectively used in personalized curriculum learning and have good effects in improving learning efficiency, providing a reference for future study on the application of the genetic algorithm in curriculum learning. However, the curriculum learning is affected by various factors; thus, only applying genetic algorithm is far from enough. Other effective methods to improve curriculum learning efficiency remain to be further studied.

3 Method and Technology

3.1 Structure of curriculum generation

Curriculum refers to the total sum of disciplines that school students should learn and the processes and arrangements. Broadly speaking, it refers to the sum of all disciplines stipulated in order to achieve the goal of school training; narrowly speaking, it refers to the knowledge system to be studied by a certain discipline, and it is a welldesigned teaching plan.

Curriculum structure refers to the form and structure of the curriculum. Figure 1 shows some examples of a professional teacher's conceptual maps related to the course of "data structure" in the computer field, which shows the relationships among courses, concepts, and learning objects. It can be concluded that the knowledge system of a course is constructed by a series of related concepts, and the learning resources of a course are composed of many learning objects. At the same time, each learning object is constructed to learn and understand the corresponding concepts. That is to say, the learning goal of the learning object is to grasp/understand/learn the concepts in the curriculum. Meanwhile, the scope of the concepts is different, that is, the granularity of the concepts is different. A concept with a larger category can contain several concepts with smaller categories. This "inclusion" relationship defines the concepts is constructed. In addition, the constraint relationship between concepts is constructed. In addition, the constraint relationship between concepts is constructed by the factor of "prior knowledge". According to the constraint relationship, the concept map can be extracted into a directed acyclic graph.

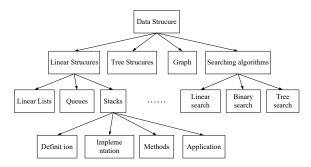


Fig. 1. An example of the structure of curriculum generation

3.2 Genetic algorithm

According to Goldberg's definition in 1986, the genetic algorithm is a kind of evolutionary algorithm and a search algorithm for solving optimization problems in computational mathematics. Evolutionary algorithms are originally developed from the phenomena in evolutionary biology, including heredity, mutation, natural selection, and hybridization. The general implementation of the genetic algorithm is computer simulation. For an optimization problem, the population of abstract representation of a

certain number of candidate solutions evolves to better decomposition, in which the candidate solution is an individual, and an abstract representation is a chromosome. The general solution is represented by 0/1 string (binary), and other expressions can also be used. The process of evolution begins with a population of completely random individuals and evolves from generation to generation thereafter. In the evolution process of each generation, the fitness of the whole population is evaluated systematically. Based on the fitness, multiple individuals are randomly selected from the current population, and new life populations are generated by natural selection and mutation as the current population in the next iteration of the algorithm. The basic flow chart of the algorithm is shown in Figure 2.

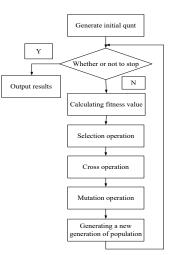


Fig. 2. The basic flow chart of the genetic algorithm

A personalized curriculum evolution model based on genetic algorithm is proposed. First, the concept coding in the curriculum generation system is used to generate the initial population by applying the intrinsic relationship in the knowledge structure diagram, and the fitness function is designed for the model. The difficulty of concepts, the relationship between learning objects and the time parameters of conceptual learning are considered comprehensively. Then, the selection, crossover, and mutation operations of the genetic algorithm are implemented, and the learning content sequence satisfying the fitness function is generated. In the whole learning process, students' learning content is updated at any time to truly meet the needs of personalized learning.

3.3 Genetic algorithm and curriculum generation

Most of the methods based on genetic algorithm apply the same process, such as analyzing students' different needs, different knowledge backgrounds and levels, generating personalized courseware sequence, and applying Horton-based courseware generation architecture, as shown in Figure 3.



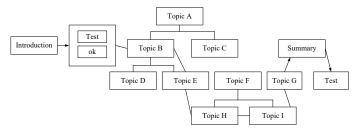


Fig. 3. Courseware generation based on Web Learning

The methods for learning path generation based on the genetic algorithm are described as follows: first, encoding to generate an initial population. If there are n courseware in the courseware database, a number (from 1 to n) is assigned to each courseware, that is to say, an integer code is used to represent a gene individually. In this way, the numbered courseware has been allocated to combine with other courseware to form a chromosome, which represents a learning path in the personalized courseware sequence; second, the initial population size is determined. Generally, the initial population size is determined by the complexity of the problem. The larger the initial population size is, the lower the computational efficiency is, but it will help to find high-quality solutions accordingly. In order to generate a more adaptive learning path, the initial population size is generally defined as 100. Third, the fitness function is selected. Fitness function is used to evaluate the fitness of the generated learning path. According to the results of the pre-test and the relationship between the difficulty coefficient and the concept of the courseware, the personalized learning path is generated. In general, the courseware mapped by learners in the items answering errors in the pre-test is considered, and the courseware with lower difficulty coefficient always chooses the first one in the learning path. The fitness function is generally:

$$f = \sum_{i=2}^{n} (1 - w) \times r_{(i-1)i} + w \times (1 - b_i)$$
(1)

In Equation (1), f is an adaptive function, $\mathbf{r}_{(i-1)i}$ denotes the degree of association between the two adjacent courseware in the generated learning path, that is, the degree of association between (i-1) courseware and the concept of the i-th courseware; \mathbf{b}_i denotes the difficulty coefficient of the i-th courseware; w denotes the adjustment weight, and n denotes the number of courseware in the personalized learning path.

3.4 System framework of personalized curriculum generation and evolution

The personalized curriculum generation and evolution system are mainly aimed at the different learning objectives, knowledge levels, and learning abilities of Internetbased network learners so as to generate appropriate curriculum knowledge structure and related learning resources for them. The subordinate, hierarchical and pioneer relationships among curriculum concepts need to be consistent with the predefined

ones in the expert system. The personalized course generation and evolution (PCGE) system framework consists of two subsystems and four functional modules. The two subsystems are personalized course generation system (PCG-LRS) based on hierarchical recommendation algorithm and personalized course evolution system based on genetic algorithm (PCE-GA) and dynamic feature analysis. The four functional modules are initial personality analysis module, learner profile description and modeling module, personality dynamic analysis module, curriculum knowledge structure management, and teaching plan formulation module. Figure 4 shows the PCGE system framework.

According to the framework of PCGE system shown in Figure 4, the workflow of the system is briefly described as follows: Step 1: Before learning, learner's user files are analyzed by the initial personality analysis module to get the initial personality data; Step 2: Teachers, based on the initial personality data of students, combined with the existing knowledge base and resources in the system, and through the module of curriculum knowledge structure management and teaching plan making, provide the knowledge structure and learning resources of the curriculum; Step 3: According to the curriculum knowledge structure and learning resource database customized by teachers, the system calls the curriculum generation algorithm, and generates the initial personalized curriculum through the initial curriculum generation subsystem PCG-LRS; Step 4: Students begin to learn the initial curriculum, and collect the dynamic knowledge structure and learning ability evaluation systematically; Step 5: In the learning process, the system read learners' dynamic personality characteristics periodically, call genetic algorithm, and generate personalized courses through PCE-GA, a subsystem of curriculum evolution; Step 6: In terms of the evolving personalized courses in the learning process of students, the system continues to collect updated data of dynamic knowledge structure and evaluation of related abilities, and repeats Step 5 until the end of the course learning.

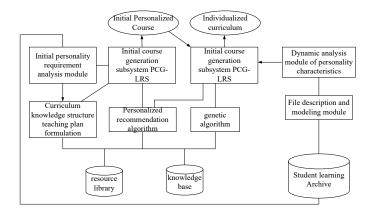


Fig. 4. System framework of PCGE

4 Results

4.1 System frame structure

On the basis of PCG-LRS, a framework of PCE-GA is proposed. There are several important problems in the process of realizing the dynamic evolution of the curriculum. Initial course generation: Based on PCG-LRS, considering the stochastic convergence of the genetic algorithm and the fast convergence of the optimal solution in the multi-objective rearrangement problem, the method of hierarchical topological ranking is introduced to avoid losing the constraints between concepts, and the generation algorithm of initial course is improved. User model updating: This is the key to PCE-GA. In order to update the user model, the system adopts evaluation-based user model learning method. Learning process evaluation: This module is the evaluation of students' learning process, which introduces vector model and is based on Q-matrixability matrix, including time parameters, difficulty coefficient, and other parameter generation or estimation. Personalized curriculum generation: The implementation process of curriculum generation algorithm based on genetic algorithm includes population initialization, fitness function, selection, crossover, mutation, etc.

The overall framework of PCE-GA is shown in Figure 5, which consists of an initial curriculum generation module, an evaluation module, a personalized curriculum generation module, and a curriculum evolution module. Initialization course (C0) generation method, as described in Chapter 3, is automatically generated through PCG-LRS according to the curriculum objectives set by teachers, the subject domain knowledge, the syllabus, and the personalized knowledge features based on user pretest results. C_0 is not the personalized course suitable for different students in the whole learning process but the overall learning content of students in a certain degree of education in a discipline. The evaluation module is to evaluate students' learning outcomes in the process of learning. Basically, the evaluation module scores a student's mastery of each concept and learning ability (0-5 points) according to the students' answers. After the concept is marked by scores in the knowledge concept map, it is called a personalized concept map (PCM), which is the main input parameter of PCE-GA algorithm. The course generated in the personalized course generation module is $C_i(i=1, 2,...,n)$ and i represents a series of courses evolved from C_0 . Throughout the process, the evaluation module evaluates students' learning performance at each set period (the cycle is set by teachers, such as weekly or biweekly). Accordingly, the curriculum C_i is evolved into C_{i+1} . In the whole learning cycle of a student learning a subject, the PCE-GA system generates a series of courses $C_i(i = 1, 2,...,n)$ until the completion of the whole subject's learning objectives.

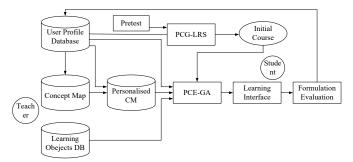


Fig. 5. The framework of personalized curriculum automatic generation and evolution (PCE-GA) based on genetic algorithm

4.2 User model update

By using user profile specification, in this algorithm, the user's knowledge and ability information are mainly updated in real time in the process of students' learning. Therefore, the updated algorithm of the user model is described in detail. The definition of related parameters in PCE-GA conceptual model and algorithm is shown in Table 1.

Symbol	Dimension	Significance	Symbol	Dimension	Significance
С	n	Concept set C= $(c_1, c_2,, c_n)$	U	s	User set U=(u ₁ ,u ₂ ,,u _s)
R	m	Resource set, that is, the set of learning objects, $R=(r_1,r_2,,r_m)$	Z	s×n	User concept U-C matrix: user's learning state matrix for con- cepts
Гij		The correlation degree of two learning objects r _i and r _j	R	m×n	The correlation matrix between m learning objects
Г	m×n	Learning object - concept R-C relational matrix	Y	s×m	User - learning object U-R evaluation matrix

Table 1. PCE-GA main parameters

User model learning is to interpret and deduce collected data, classify noise, form useful knowledge about user characteristics, and format user model with final structured representation. Common centralized learning technologies include TF-IDF (Term Frequency--Inverse Document Frequency), Bayesian classifier, decision tree induction, neural network, and clustering. However, in this system, the user's personality characteristics are not the traditional user interest characteristics, but the user's grasp of each concept, learning ability, and different target characteristics in the knowledge system in the course of learning curriculum. Therefore, the above learning methods are not available. As a result, an evaluation-based user model learning method is proposed to obtain a user's personality characteristics implicitly based on the evaluation of knowledge and ability in the process of user learning.

The automatic curriculum generation system is designed to enable professional teachers to independently complete the process of curriculum construction and personalized curriculum generation without the guidance of professional technicians, and try not to increase the burden of professional teachers so that they can cope with the teaching process freely in the face of large-scale learning student groups. Therefore, in the system adopted, all the learning objects and self-test questions designed and developed by teachers are regarded as curriculum resources that can be used to evaluate students' learning status, i.e. curriculum resources that can be assessed, without the need for a large number of specialized testing problems as described in Birenbaum. Based on the above principles and methods, Q matrix is extended to "concepts-learning objects" correlation matrix, and a series of different types of "assessable teaching resources" are designed to provide a formative evaluation in the learning process.

Drawing on the idea of formative evaluation in pedagogy, in the course of designing learning resources, professional teachers define evaluation methods according to metadata criteria of learning objects, including exercises and tests related to learning resources, or subjective evaluation methods that can determine the scores, etc. Therefore, in the learning process, the system can get students' learning state of each learning object at any time. In this system, the evaluation process generates a user-learning object evaluation matrix UR (User Resource), in which the value of each component is 0-1/NA, representing the score value (0-1) or the unlearned value, respectively. The UR form is shown in Table 2.

Vmatuiv	Learning resource R				
Y matrix	r 1	r 2		r _m	
	u ₁	y11	y ₁₂		y _{1m}
The second t	u ₂	y ₂₁	y ₂₂		y _{2m}
User U					
	us	y _{s1}	y _{s2}		ysm

Table 2. User learning object evaluation matrix UR definition form

4.3 Experimental results

Because of the role of selection operator and crossover operator in genetic algorithm, some excellent gene segments are lost prematurely, which limits the search scope, so that the algorithm can only find the local optimal solutioninstead of the global optimal solution. In order to overcome the problem of premature convergence, relatively excellent parameter combinations are determined through experiments. These parameters include population size, crossover probability P_c , mutation probability P_m , selection crossover and mutation, and other basic operators. Control parameters are important to system performance. For example, if the population size is too small, the search range of the algorithm will be affected, leading to the failure to obtain the global optimal solution; if the population size is too large, the search efficiency of the algorithm will be reduced. The bigger the crossover probability and mutation probability are, the stronger the detection ability of the algorithm is, but the easier it is to destroy the stability of the algorithm, so as to detect new hyperplanes; the smaller the crossover and mutation probability are, the stronger the development ability of the algorithm is, and the better the suboptimal individuals can be maintained in the algorithm. However, in the application environment of curriculum generation, it is necessary to determine the optimal combination of parameters through experiments according to the actual learning data.

In the experiment, the initial population is set to 50, and repeated experiments are conducted to determine the crossover probability and mutation probability, so as to achieve good results. First, the crossover probability is set to 0.5, and the mutation probability is set to different values to observe its effects on the average fitness protocol and the optimal adaptive chromosome generation. Figures 6 and 7 show the results of the tests and comparisons. The test results show that the variation probability has a small effect on the average fitness of the population, and a high enough average fitness will change the evolutionary trend of the population.

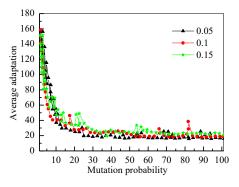


Fig. 6. The effect of mutation probability on the average fitness

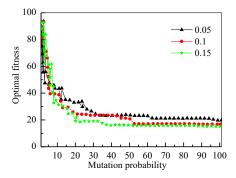


Fig. 7. The influence of mutation probability on optimal fitness

Next, the mutation probability is set to 0.1, while the crossover probability is set to different values to observe its effects on the average fitness protocol and optimal adaptive chromosome generation. The experiments shown in Figures 8 and 9 show that higher crossover probability and mutation probability can avoid premature popu-

lation convergence. However, the emergence of a large number of new chromosomes means that the performance of population evolution is bound to decline with the super-large-scale calculation. Considering many tests, the crossover probability is set to 0.5 and the mutation probability is set to 0.15.

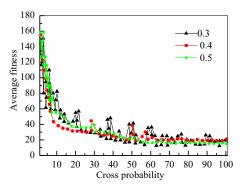


Fig. 8. The influence of crossover probability on the average fitness

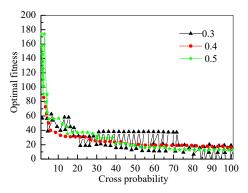


Fig. 9. The influence of crossover probability on optimal fitness

For the determination of selection operator, the classical roulette wheel selection method, tournament selection method, and order statistic based competitive selection method are compared. After repeated experimental tests, the tournament selection method based on fitness ranking order statistics is chosen. Based on the principle of avoiding damaging the excellent gene pattern in the acquired parent generation as much as possible, considering that all genes in chromosome must appear and not repeat, a partial matching crossover algorithm (PMX) similar to double crossover is chosen, and a mapping relationship is designed according to the practical problems of course content arrangement.

5 Conclusion

The personalized realization of curriculum learning in the network learning environment is focused on the generation and evolution model of curriculum content and its algorithm implementation in the formal learning process. The problem of individualization of learning content in the process of large-scale network education is explored and solved. Combining with the genetic algorithm, the automatic generation of curriculum knowledge structure and learning objects in the learning process is realized, and the learning content is updated dynamically in the learning process.

Through the successful application of the genetic algorithm in the whole learning process to realize the dynamic updating of the curriculum, a personalized curriculum evolution algorithm based on genetic algorithm PCE-GA is proposed. In order to quickly converge to an optimal solution in multi-objective problems, a hierarchical topological sorting algorithm is proposed to generate the initial solution space considering the directed acyclicity of concept maps and combining with the topological sorting algorithm. Under the framework of this algorithm, the dynamic updating of the target user model and the continuous evolution of the curriculum provide a good framework with good versatility and expansibility for solving the personalized curriculum generation problems in the network learning environment.

6 Acknowledgement

This work was supported in part by the key platform and research project (innovative strong school project) of the Guangdong Provincial Department of Education in 2016; Research and Implementation of the Credit System Reform Program in Colleges and Universities-Taking the China Business School of Guangdong University of Finance and Economics as an example (Number: 2016WQNCX209).

7 References

- [1] Girgis, M. R., Ghiduk, A. S., Abdelkawy, E. H., Girgis, M. R., Ghiduk, A. S., & Abdelkawy, E. H. Automatic generation of data flow test paths using a genetic algorithm. International Journal of Computer Applications, 2014, vol. 89(12), pp. 29-36. <u>https://doi.org/10.5120/15684-4534</u>
- [2] Yin, Y., Thapliya, R., & Zimmermann, R. Encoded semantic tree for automatic user profiling applied to personalized video summarization. IEEE Transactions on Circuits & Systems for Video Technology, 2018, vol. 28(1), pp. 181-192. https://doi.org/10.1109/tcsvt.2016.2602832
- [3] Chen, J. W., Yang, H. J., Xu, N., & Li, L. Scheme automatic generation based on full life cycle assessment gene model of structure. Applied Mechanics & Materials, 2014, vol. 687-691, pp. 1622-1627. https://doi.org/10.4028/www.scientific.net/amm.687-691.1622
- [4] Shi, X., & Zhang, B. (2014). An automatic generating method of 3d personalized clothing prototype, 2014, vol. 13(12), pp. 127-137.

- [5] Yu, L., Man, L., & De-Qi, X. Research on the application and improvement of automatic test paper generation based on genetic algorithm. Modern Computer, 2017, vol. 30(3), pp. 22-29.
- [6] De, S., Bhattacharyya, S., & Dutta, P. Automatic magnetic resonance image segmentation by fuzzy intercluster hostility index based genetic algorithm: an application. Applied Soft Computing, 2016, vol. 47, pp. 669-683. <u>https://doi.org/10.1016/j.asoc.2016.05.042</u>
- [7] Sahu, R. K., Panda, S., & Padhan, S. Optimal gravitational search algorithm for automatic generation control of interconnected power systems. Ain Shams Engineering Journal, 2014, vol. 5(3), pp. 721-733. <u>https://doi.org/10.1016/j.asej.2014.02.004</u>
- [8] Xin-Huai, Q. U., Sheng-Bo, H. E., Bi-Rong, D., & Chang, X. U. Improved differential evolution cellular genetic algorithm for product configuration under mass customization. Modular Machine Tool & Automatic Manufacturing Technique, 2017, vol. 65(1), pp. 46-65.

8 Author

Xiaocong Duan Huashang college, Guangdong University of Finance & Economics, The research content, genetic algorithm, personalized learning, curriculum generation, curriculum evolution (feyhil@163.com).

Article submitted 2019-05-07. Resubmitted 2019-05-26. Final acceptance 2019-05-27. Final version published as submitted by the authors