Curriculum optimization algorithm using genetic programming methods

Curriculum optimization algorithm using genetic programming methods

A K Guts[1], A A Senkovskaya[2] and I I Furayeva[2]

[1]Department of Computer Science, Dostoevsky Omsk State University, 55a Mir avenue Omsk, Russia

[2]Department of Information technologies, L.N. Gumilyov Eurasian National University, 2 Satbayev st. Nur-Sultan, The Republic of Kazakhstan

Abstract. Recently, optimization has become an important aspect of educational process management. The goal of the optimization process is usually reduced to reducing the total amount of the pedagogical load of the university, as well as to the formation of an optimal schedule of training sessions with the existing classroom fund. This article discusses the main aspects of the formation of curricula in order to optimize the educational process. The problem of formation of educational flows is considered with a combination of disciplines of general education and majors. A solution is proposed to optimize the curriculum in order to minimize educational flows and the total teaching load of departments using genetic programming methods.

1. Introduction

The formation of the optimal load of the university is one of the most important tasks of planning and managing the educational process. The initial document for determining the initial academic load, in turn, is a modular educational program and a curriculum based on it for the corresponding year of study.

In the current conditions of the organization of the educational processes of the university, optimization becomes an integral part of the effective management of the university as a whole. As a rule, optimization processes are carried out over the curriculum in order to minimize the number of training flows, and therefore, to minimize the teaching load of the departments.

Consider some of the factors that influence the process of formation of the teaching load.

The number of credits performed by the teacher at one rate depends not only on the number of students, but also on the total load of the university. Achieving the uniformity of the training load by semester at each department and its competent distribution make it possible to create an optimal training schedule from the position of students (namely, without windows). The uniformity of the load at the department is understood as its sufficient, but not absolute uniformity, given the fact that the spring semester is planned for training and production practice, and also there are no classroom studies at the final courses of undergraduate and graduate courses.

The time standards for the formation of the pedagogical load are developed by the university in order to reduce the total load of the department and the university as a whole, and therefore to reduce the load on one teacher's rate. Since the main contingent of students, as a rule, are students in undergraduate programs, the most effective is the optimization of undergraduate curricula. It is only possible to solve the problem of optimizing in-line lectures on the disciplines of the general educational component taught for all specialties of the faculty and the university to minimize the overall load of the university with the approved time standards for all types of classes.

At the 3rd and 4th year, unlike the younger courses, classes are held by specialty, and mainly by the graduating department. Freshmen study general education subjects, having no more than one discipline per semester in their specialty. Only in the 2nd year of undergraduate courses are provided as disciplines of the general educational component, which are read by the service departments of the university, and a sufficient number of disciplines of the basic and core components taught by the graduating departments.

Each university, based on the availability of the necessary classrooms for in-class lectures, independently determines the maximum number of students in one stream for conducting such classes, both for the cycle of general subjects and for basic/specialized disciplines. These standards are the basic indicator that allows you to optimize the load of the university by increasing the number of in-line lectures, with the number of students approaching the maximum.

2. Statement of the problem curriculum optimization

Consider the second year undergraduate curricula of all faculties and specialties of the university. Curricula, or educational programs, are approved prior to student enrolment.

Each curriculum contains disciplines of the general educational component, which are read by the service departments of each faculty, for example, such as Philosophy, Sociology, Computer Science, Higher Mathematics and others. Some of these disciplines are compulsory for all university students (for example, Philosophy, Political Science), some are chosen by students of each specialty (for example, Entrepreneurship and Business, Professionally Oriented Languages by profession). In addition, each faculty has the same disciplines for most or all specialties (Physics for all technical specialties, Microeconomics/Macroeconomics for economic specialties, General and social psychology for social specialties and others). In addition to general educational disciplines, for second-year students, in contrast to the first year, the largest number of disciplines unique to each specialty is already provided.

It should be noted that in Kazakhstan, the enrolment of students is carried out in the Kazakh and Russian departments. Moreover, the number of budget places allocated by the Ministry of Education and Science varies significantly for different areas of preparation. At the same time, the enrolment of students varies in a significant range, for example, from 186 students in the Kazakh department and 46 students in the Russian department with a degree in Computer Engineering and Software, up to 7 students and 3 students in the Kazakh and Russian departments, respectively, with a degree in Turkology. It should be noted that over the past few years, the popularity of the specialties of the Faculty of International Relations and the Faculty of Law has significantly decreased, in turn, there has been an increase in interest in the specialties of the Faculty of Information Technology: the total set of faculties for the last academic year is 198/143/949, respectively.

However, it should be noted that the number of small groups over the past 5 years has increased even in large universities (again, due to the popularity of a particular group of specialties). The developed educational programs used to form the training load require analysis taking into account the real student population.

Consider the situation in a university that provides instruction in two languages. In the academic year 2018-2019, at 13 faculties, you can get 1/4 of the Kazakh students and 1/3 of the Russian students studying the subjects of the general education component in one semester, and 3/4 and 2/3 of the students in the other semester, respectively, which creates significant imbalance in the load of service departments. The current situation significantly complicates the possibility of scheduling training sessions, up to the complete impossibility of obtaining the length of the schedule within one shift. In addition, the overall workload of teachers with numerous windows in the schedule is increasing. Thus, the optimization of the teaching load becomes relevant for both small universities and large universities (from the standpoint of student contingent and the variability of the offered training specialties). Moreover, the factor should be taken into account that the general pedagogical load of the university is determined not only by the curriculum, but also by the contingent of students.

Due to the fullness of the curriculum with disciplines of a certain component, the analysis with the aim of optimizing the academic load can be divided into 3 main parts:

A.   analysis of the curricula of the first course of study - the main part is made up of disciplines of the general educational component that form the load of service departments [1];

B.   analysis of the curricula of the third course of study - consists of disciplines of the main component taught by graduate departments;

C.   analysis of the curricula of the second course of study - consists of both disciplines of the general educational component that require optimization of learning flows, and disciplines of the basic / profiling component that forms the burden of graduating departments.

It makes no sense to analyze the graduate curriculum, since theoretical training is carried out only in the fall semester, and passive types of work are planned in the spring semester (production practice, writing a thesis).

Consider the statement of the problem of optimizing the curricula of the second course of study in order to minimize the academic load of the university. The disciplines of second-year students of each department can be conditionally divided into 3 categories: general university (usually general education, taught for all specialties of the university), general faculty (basic disciplines that characterize the profile of the faculty) and department (specialized disciplines of graduating departments). It should also be noted the fact that students of several specialties or terms of study can study at the department.

Nevertheless, in contrast to the task of analysing the first year curriculum, when the number of students is planned and considered preliminary, the contingent of students is given in the task of optimizing the second year curriculum. Then by editing the curriculum, i.e. transferring disciplines not related to prerequisites to another semester with maintaining the total number of credits in each semester in all curricula, a solution to the problem of minimizing the teaching load by forming lecture flows that are as close as possible to the approved upper values of the flow fullness range, as well as optimizing the load of each department.

We examine in detail each of the levels of disciplines. The placement of university-wide disciplines in the curriculum is most easily edited, since the overall reduction in load meets only a positive reaction from the leadership of the university, and the consolidation of disciplines for service departments does not change. The level of inter-faculty interaction is more complex. Even the agreements reached between the departments at the stage of formation of educational programs are often violated by the beginning of the school year. But be that as it may, by the beginning of the analysis of the second year curricula, the final consolidation of disciplines in the departments and the student contingent studying according to each of the analysed curricula are known.

It would not be desirable in this case to apply the greedy algorithm to the solution of the optimization problem, that is, first optimize the disciplines of the first level, then the second and only then the third level, but in the general case it was not possible to prove that the problem under study is described by a matroid.

Of course, if in all the curricula of the second year of study the number of credits in each semester for disciplines of the 1st level was the same, then the solution of the problem for disciplines of the 2nd and 3rd levels would not depend on the optimization of the distribution of disciplines of the 1st level. It should be noted that the second level may not exist for all faculties, but a third level is available for each department.

Let us consider in more detail the formation of flows for an individual discipline of the first level. As a zero approximation, we will use the original curricula.

Figure 1 shows an example of the selection of flows using the capabilities of MS Excel, which is often performed in the manual training schedule department using the SUMPRODUCT function.

As can be seen from the figure 1, flows can be formed from students studying in the specialties of different faculties. With the maximum occupancy of one stream in the discipline of 120 students, 15 flows were formed in the Kazakh branch, in the Russian branch - 8 flows.

When using the program, this rather long and responsible work is performed automatically. The resulting zero approximation is checked for optimality. The optimization task is not to divide the total number of students into 2 parts, with approximately the same amount, i.e. distribute each discipline equally for 2 semesters.

The task is formulated more difficult: to divide the student body of faculties in such a way that in each semester the number of students is approximately a multiple of the maximum flow rate (in this case 120 students) regardless of the language of instruction (in the conditions under consideration for the Russian and Kazakh departments). In this case, flows are formed as shown in figure 1.

3. Modelling the process of optimization of curricula

The stated problem is non-trivial and complex, and the search for the optimal solution requires certain costs. This task can be attributed to planning tasks and multi-criteria tasks requiring a multi-stage

solution. To solve the formulated problem of optimizing the curricula of the second year of study, a non-mathematical approach can be used, namely, a solution using genetic algorithms [2].

Faculty	Name of specialty	Kazakh language of training		Russian language of training		241	225	107	120	107	209	120	117	102	239	119	221	241	230	227
		Students	Groups	Students	Groups	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4
Faculty of Architecture and Civil Engineering	Surveying and cartography	62	3	29	2	1											1			
	Design (up to 5 years)	35	2	18	1		1										1			
	Architecture (up to 5 years)	36	2	26	1		1										1			
	Fine Art and Drawing	2	1	0	0		1										1			
	Construction and development	179	6	41	2	1											1			
	Engineering systems and networks	130	5	13	1		1											1		
	Total faculty:	**444**		**127**																
Faculty of Transport and Energy	Transport, transport equipment and technologies	70	3	17	1				1									1		
	Heat power engineering	107	4	17	1			1										1		
	Standardization and certification	50	2	8	1					1								1		
	Organization of transportation, movement and operation of transport	45	2	11	1						1							1		
	Total faculty:	**272**		**53**																
Physics and Technology Faculty	Nuclear physics (up to 5 years)	62	5	11	1						1							1		
	Technical Physics	22	1	6	1	1												1		
	Radio engineering, electronics and telecommunications	192	7	24	1								1					1		
	Physics	17	1	4	1								1					1		
	Physics (Natural Sciences)	29	1	4	1									1				1		
	Space Engineering and Technology	47	2	8	1									1				1		
	Total faculty:	**369**		**57**																

Figure 1. Fragment table selection flows

Consider the issue of flow optimization for disciplines of the general educational component, i.e. general university level. After looking at all the curricula of the second year of study, two such disciplines were generally identified. In the existing curricula, they can be located both in different semesters, and in one semester of different specialties. In addition, if one discipline - Philosophy - is studied by all students, then the second discipline is not required, if it was possible to put our own core discipline in the educational program of the specialty. Therefore, the issue of the formation of continuous lectures can be decided independently for each discipline. Suppose that the discipline is presented in the curriculum in such a way that it is studied simultaneously for each faculty.

We will form a contingent of students for each faculty, taking into account bilingual education in the form of a table, separately for the fall and spring semesters. At the next stage, it is necessary to determine the total number of students to study the disciplines in question and, accordingly, the minimum number of flows for each language of instruction, a multiple of the maximum fill rate.

Suppose that with a maximum occupancy of a multiple of 120 students, the size of the flow should not actually exceed 123 students, taking into account a 2.5% deviation in the formation of the flow. This allows you to correctly round the resulting value of the number of flows.

The task of flow optimization for the disciplines of the general university component has the peculiarity that the global minimum is known, but it may not be achievable.

Consider the flow generation algorithm using a specific example. The peculiarity of the algorithm is that in addition to determining the number of flows, the streams themselves are also obtained as an additional result.

Data from the curriculum for the fall semester is presented in figure 2.

After the student contingent has been formed to study the general educational disciplines of the fall semester, it is possible to determine whether it is possible to create associations from groups of faculties that allow you to get a whole number of flows, with priority being given to the language of instruction with the largest number of students. It is easy to determine that such a distribution has the form as shown in figure 3.

The table in figure 3 shows the total number of flows and the number of students. Based on the data presented, no faculty alone was able to form an integer number of flows. Let us consider in more detail the formation of individual flows. In addition, it is important to note that at this stage, flows are formed from different specialties of various faculties.

Kazakh language of training	Russian language of training	Faculty
233	29	Faculty of Mechanics and Mathematics
249	78	Faculty of Social Sciences
30	13	History Faculty
132	66	Faculty of International Relations
104	39	Faculty of Law
773	176	Faculty of Information Technology
506	107	Faculty of Natural Sciences
128	37	Faculty of Journalism and Political Science
2155	545	
17.96	4.54	

Figure2. The formation of the student contingent in faculties for the fall semester

As can be seen from figure 3, the number of students per flow does not exceed the maximum fill rate, taking into account 2.5% deviation (i.e. 123 students). It is important to note that the formation of flows manually led to the unification of the groups into 26 flows of the Kazakh branch and 10 flows of the Russian branch, and a theoretical minimum of 18 and 5 flows was achieved by software.

			KZ				RU		
			12.01	2.08	1.94	1.93	2.02	0.65	1.88
			1	2	3	4	1	2	3
KZ	RU	Faculty	1441	249	233	232	242	78	225
233	29	Faculty of Mechanics and Mathematics			1		1		
249	78	Faculty of Social Sciences		1				1	
30	13	History Faculty	1						1
132	66	Faculty of International Relations	1						1
104	39	Faculty of Law				1			1
773	176	Faculty of Information Technology	1					1	
506	107	Faculty of Natural Sciences	1						1
128	37	Faculty of Journalism and Political Science				1	1		

Figure 3. The formation of educational flows for a given student contingent

Figure 4 shows that students of 2 faculties and 15 groups are selected to form three flows.

			KZ		RU		2.94
							353
Faculty	Specialty code	Name of specialty	Students	Groups	Students	Groups	1
Faculty of International Relations	5B050500	Regional Studies	25	1	6	1	1
Faculty of Natural Sciences	5B060900	Geography	25	1	2	1	1
	5B061000	Hydrology	41	2	8	1	1
	5B060600	Chemistry	12	1	1	1	1
	5B070100	Biotechnology	250	10	54	2	1

Figure 4. Example of thread formation

Let us consider in more detail the search algorithm for the optimal formation of flows for one discipline.

1. Determine the total number of students in each department.

2. Determine the minimum integer number of flows for each branch per year.

3. Determine the minimum number of threads per semester according to the current state of the curriculum.

4. If this number is an integer, then the solution is considered acceptable, otherwise a genetic algorithm is used to determine the smallest change in the curriculum that will lead to integer results.

If the procedure for applying the genetic algorithm is standard, and the chromosome is considered only for the fall semester, and the number of elements is equal to the number of faculties, then the task of evaluating the chromosome is much more complicated than at other levels. Moreover, it is important to consider that the assessment must be carried out not only for the autumn, but also for the spring semester.

- Choose the faculty with the largest number of students.
- Choose for him the following faculties, allowing to obtain a total number of students close to a multiple of 120.

A greedy algorithm for this task can give an upper bound on the bound. The sum of such boundaries for both semesters will give an estimate of the chromosome. To form specific flows, the described algorithm is applied to each group of faculties, and individual specialties are considered.

For a more accurate result, the problem of forming flows by specialties should be solved, if it is considered acceptable that the faculty specialties study the discipline in different semesters.

If the second discipline is also studied by all university students, then the resulting distribution is applied to the second discipline.

After the formation of the smallest number of continuous lectures, the curriculum system will be generally removed from the state of equilibrium. You can try to compensate for it with the help of the second discipline, or you can solve this issue at a lower level, accepting the resulting distribution as an initial approximation.

Let us evaluate the contribution of each task level to reduce the total number of threads. At the first university level there are 2 disciplines. At the second, general faculty level, there is at least one discipline that all students of the faculty study. The division of such discipline into departments is less significant for large sets of students for each department and more significant for specialties from small student populations. In the worst case, the decrease in flows can reach a number comparable to the number of faculties. At the department level disciplines of one department are considered. Stream optimization with 2 specialties and 2 forms of training (full and accelerated training period) can reduce the number of lectures to three from each discipline. After one pass of the algorithm, iterations are repeated until the required level of accuracy is achieved.

Thus, not resolving the collisions at each level, but transferring their solution to a lower level, and iterating through the levels, we obtain a solution to the problem of stream optimization for the entire university [3].

For the problem under study, a consistent application of a series of genetic algorithms, which can be called the Comb method (Figure 5), was proposed. The upper part of the so-called ridge covers and connects all the disciplines of the general educational component, presented at the university level, the middle part - the general disciplines for each faculty (if any), and the lower part as separate cloves - profile disciplines of the department.

In some cases, as described earlier, faculty level disciplines can be replaced in the curriculum with specialized departmental disciplines. Then the work of combining flows at the faculty is not carried out, but the transition to the analysis of cathedral flows is carried out. Such a situation in figure 5 is represented by a block with dashed borders.

As an implementation of the search for the optimal solution at each level of the comb, an island model of the genetic algorithm was chosen.

The island model is one of the variants of the parallel genetic algorithm model. The essence of the island model of the genetic algorithm is to divide the general set of individuals into equal groups of subpopulations, usually of a fixed size. Applying a certain genetic algorithm during the execution of processes, each of these subpopulations develops separately from each other. Therefore, we can conclude that the individuals settled on different islands.
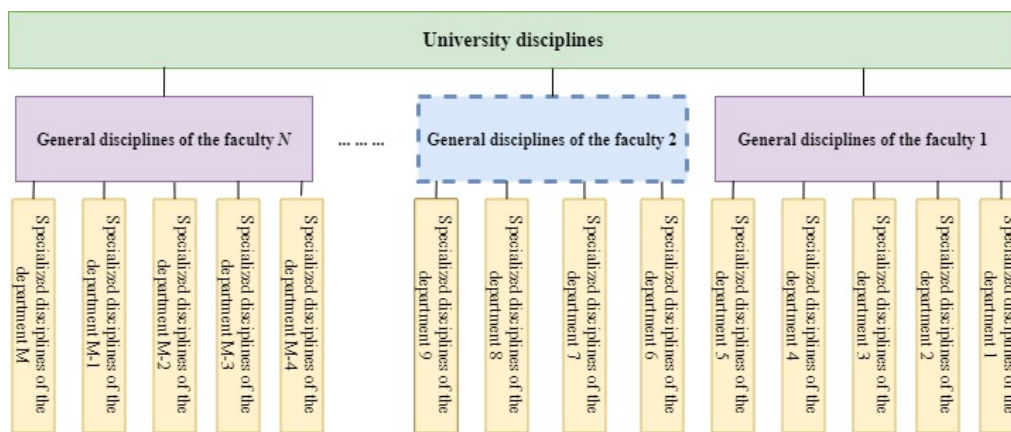
Figure 5. Schematic representation of the Comb method

Occasionally, for example, every 5 generations, islands (i.e. processes) exchange several good individuals. This operation is called migration, which allows process islands to exchange genetic material.

A distinctive feature of the island model is that each sub-population can use different strategies for the selection and formation of a new generation. This factor allows you to get the most diverse solution options on various process islands.

Since genetic algorithms are stochastic in nature, with different launches, the same population can converge to different solutions. Therefore, the use of the island model allows us to implement the algorithm several times at once, which will allow us to combine the "achievements" of different islands to obtain the optimal solution in one of the subpopulations [4].

Application of the proposed Comb method allows us to solve the problem of optimizing curricula for the second year of study, taking into account several criteria: minimizing the flow of training while observing the acceptable distribution of the teaching load of departments and faculties in general, as well as the minimum number of movements in the analyzed curricula.

4. Conclusion

In the course of modeling the process of optimizing the curricula of the second year of study, the Comb method was proposed, which, when using one of the options for parallel genetic programming, allows us to solve the problem of minimizing educational flows.

The proposed method differs from the ordinary island model of the genetic algorithm in that it allows you to search for a solution, going down sequentially from the upper level (general educational disciplines) to the faculty and further to the cathedral levels. Thus, the search for the optimal solution (i.e., minimization of study flows) can be performed locally first within the framework of one faculty, and then, taking into account the previous stage of optimization, an analysis of the next faculty is carried out.

Ultimately, the proposed options for optimizing the curriculum are considered by the staff of the curriculum, and then a decision is made, agreed with the relevant departments, on making changes to the curricula of the second course of study.

5. References

[1]   Senkovskaya A and Furayeva I 2019 *Analysis of the source data in the task of optimizing work curricula* (Mathematical Structures and Modeling, vol.2 (50)) (Omsk: Dostoevsky Omsk State University, Russia) pp.77-86.

[2]   Stevens R 2016 *Algorithms. Theory and practical application* (Moscow: Eksmo).

[3]   Tokarev V 2012 *Methods of optimal solutions. In 2 vols. Volume 2. Multicriteria. Dynamics. Uncertainty* (Moscow: Fizmatlit).

[4]   Whitley D 1994 *A genetic algorithm tutorial (Statistics and Computing, vol.4, issue 2)* pp.65-85.