Research and analysis of course arrangement based on genetic algorithm

Guangmin Sun \textsuperscript{a} Yahui Li \textsuperscript{b}

The Information Department, Beijing University of Technology, Beijing 100124, China

\textsuperscript{a}gmsun@bjut.edu.cn, \textsuperscript{b}137623637@qq.com

Abstract. Gene coding has always been an important and difficult problem in the course scheduling of genetic algorithm. Starting from the definition of the set and the limited conditions of the scheduling problem, this paper expounds the coding method of the scheduling problem of genetic algorithm in detail. The gene coding method in this paper is a decimal coding with special significance. The information of each course: teaching class, class time, course name, teaching teacher, classroom type and other information are encoded as chromosomes, and the coding method is more standardized and accurate. The crossover rate and variation rate of the designed genetic algorithm will automatically change with the change of fitness, and the self-adapting crossover rate and variation rate can make the algorithm fleetly get the globing greatest result. The fitness function is improved, and the traditional genetic algorithm is compared with the improved genetic algorithm. From the final results, It shows that the improved genetic algorithm is more excellent in efficiency and performance.

Keywords: Decimal coding, adaptive change, fitness function.

1. Introduction
The arrangement of curriculum in Colleges and universities is the most basic and very important teaching management work in Colleges and universities. This ensures the stable establishment of school teaching order, and also reflects the training of qualified talents and the specific implementation of education policy. Usually, after the task arrangement of academic affairs is made, The academic affairs office receives the work report, which leads in the teaching schedule and randomly arranges the courses. How to arrange the classroom, teachers, students, curriculum and time reasonably through the constraints is the key to solve the problem [1]. The ultimate goal of course arrangement is to arrange teaching plan reasonably in time and space.

2. Definition of Set
The resource allocation model is a simplified form of the mathematical model of the course arrangement problem, that is, under specific constraints, different needs of different individuals are allocated corresponding resources [2]. Our task is to arrange classes, teachers, courses and classrooms according to the teaching plan in a certain time period of the week without conflict. Therefore, Related sets are defined as follows:.
• Room set: R \{r_1, r_2, r_3, ..., r_n\}. The basic content of each classroom includes the name, property, number, etc.
• Time set: T \{t_1, t_2, t_3, ..., t_n\}. Each element T represents a time segment.
• Class set: C \{c_1, c_2, c_3, ..., c_n\}. The collection of classes to be arranged, in which the elements are classes. The basic contents of each element include department, major, class name, etc.
• Person set corresponding to class: P \{p_1, p_2, p_3, ..., p_n\}.
• Lesson set: L \{l_1, l_2, l_3, ..., l_n\}. Each course includes course type, number of classes.
• Mentor set: M \{m_1, m_2, m_3, ..., m_n\}. Basic information of each teacher includes number, name, Department, etc.

3. Classification and Design of Constraints
According to the analysis of the problem of arranging courses, the limited conditions can be divided into hard limited conditions and soft limited conditions.

3.1. Hard Constraints
Hard constraints are the conditions that must be met in the course of solving the scheduling problem, and they also have the characteristics of objective logic [3]. We design hard constraints from the perspective of class and classroom.
• No more than one course can be taught in a class in a certain period of time.
\[ \forall l, t \mid (\exists c(l, t_k) = c(l, t_j) \land l \neq t_j)), t_k \in T, t_j \in L, l_j \in L \]  (1)
Where: \( c(l, t) \) is the class of \( l \) course in \( t \) period.
• The capacity of the classroom is larger than the number of students who need to take classes in the classroom.
\[ \forall r_i \in R, \forall t_k \in T, \text{Capacity}(r_i) \geq P(c_j) \]  (2)
Where: \( \text{Capacity}(r_i) \) represents the capacity of \( r_i \) classroom, \( P(c_j) \) represents the number of \( c_j \) class.

3.2. Soft Constraints
Soft constraints reflect the idea of human text. Although they are not required to be met in the scheduling problem, but these soft limited conditions should be up to the standard as far as possible, so that a more reasonable timetable can be generated.
• The courses in the student curriculum shall be evenly distributed as far as possible, and the same courses shall be arranged in different periods.
\[ \forall p_i \in P, \forall l_j \in L, \text{Min}\left\{ \sum_{t \neq t_i}^{n} (P(t) - P(t_i))^2 \right\} \]  (3)
Where \( P(t) \) is the time of class.
• As far as possible, courses with small number of students should not be arranged in classrooms with large space.
\[ \forall r_i \in R, \forall c_j \in C, \text{Max}\left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P(c_j)}{\text{Capacity}(r_i)} \right\} \]  (4)
Where: \( \text{Capacity}(r_i) \) represents the capacity of \( r_i \) classroom, \( P(c_j) \) represents the number of \( c_j \) class.

4. Coding Method
In this improved genetic algorithm, the specific information of each course is encoded by decimal gene coding. The chromosome code contains the information of teaching class, class time, course name, teacher, class type, etc. There are 30 codes in total.
Table 1. Schedule Gene Coding

| Class code | Class hour code | Course code | Teacher code | Classroom code |
|------------|----------------|-------------|--------------|----------------|
| 7-digit decimal number | 5-digit decimal number | 7-digit decimal number | 6-digit decimal number | 5-digit decimal number |

- The class code includes the year of enrollment, the Department, the major and the class number, which are expressed in 2-digit, 2-digit, 2-digit and 1-digit decimal system respectively.

Table 2. Class Code

| Enrollment Year | Affiliated departments | Major | Class number |
|-----------------|------------------------|-------|--------------|
| 0               | 0                      | 0     | 1            |
| 9               | 9                      | 9     | 9            |

- Class hour code refers to the code of class time, including how many weeks, what day and section, which are expressed in 2-digit, 1-digit and 2-digit decimal systems respectively.

Table 3. Class Hour Code

| How many weeks | What day | Section |
|----------------|---------|---------|
| 0              | 0       | 0       |
| 1              | 7       | 9       |

- The course code includes course type, number of classes, number of classes per week and course number, which are expressed in decimal system of 1 digit, 1 digit, 1 digit and 4 digits respectively.

Table 4. Course Code

| Course type | Number of classes | Weekly classes | Course number |
|-------------|-------------------|----------------|---------------|
| 1           | 0                 | 0              | 0             |
| 2           | 9                 | 9              | 9             |

- The teacher code includes the Department, the title and the number of the teacher, which are expressed in two decimal places, one decimal place and three decimal places respectively.

Table 5. Teacher Code

| Affiliated departments | Title of teacher | Teacher number |
|------------------------|------------------|----------------|
| 0                      | 1                | 0              |
| 9                      | 9                | 9              |

- The classroom code includes the building, floor and door number, which are represented in 1-digit, 2-digit and 2-digit decimal system respectively.

Table 6. Classroom Code

| Building | Floor | House number |
|----------|-------|--------------|
| 1        | 0     | 1            |
| 9        | 2     | 9            |

5. Adaptive Crossover and Mutation Rates
The selection range of crossover probability $p_c$ of traditional genetic algorithm is 0.40-0.99[6]. The selection range of mutation probability $p_m$ is 0.001-0.1[4]. However, according to the characteristics of the increase of iterations, the rate of crossover and the rate of variation cannot remain the same. The crossover rate and mutation rate of the genetic algorithm used in this paper will change spontaneously.
according to the change of fitness. For the individual with higher adaptive value than the average, there is a lower cross rate and variation rate, it's not easy to be eliminated in the next generation; for the individual with lower adaptive value than the average adaptive value of the population, there is a relatively higher cross rate and variation rate, this makes it easier to remove. Therefore, the genetic algorithm of adaptive change crossover and mutation operation used in this paper is easier to converge to the global optimal solution.

The mathematical formula for the cross rate \( p_c \) and variation rate \( p_m \) of adaptive change is as follows.

\[
P_c = \begin{cases} 
P_c(f_{\text{max}} - f)/(f_{\text{max}} - f_{\text{avg}}), & f \geq f_{\text{avg}} \\
P_c, & f < f_{\text{avg}} 
\end{cases} \tag{5}
\]

\[
P_m = \begin{cases} 
P_m(f_{\text{max}} - f)/(f_{\text{max}} - f_{\text{avg}}), & f \geq f_{\text{avg}} \\
P_m, & f < f_{\text{avg}} 
\end{cases} \tag{6}
\]

Among them, \( f_{\text{max}} \) represents the maximum adaptive value of the population, \( f_{\text{avg}} \) represents the average adaptive value of the population, \( f \) is the larger fitness value of the two individuals to be cross operated, \( f' \) is the fitness value of the individual to be mutated, \( p_c \) and \( p_m \) is a constant.

### 6. Fitness Function Design

In this paper, the improved genetic algorithm is based on the traditional genetic algorithm. In this paper, when analyzing the design of course scheduling problem, the solution of resource conflict and the search for the optimal solution of combinatorial planning are affected by the adaptability of the algorithm. A variety of restrictions affect the result of course arrangement. The corresponding functions are designed by these restrictions. The comprehensive result of each corresponding function is the fitness function of the algorithm.

\[
f_g = \frac{1}{1 + crash_g} + reward_g + \sum_{j=1}^{n_{\text{major}}}(\sum_{i=1}^{n_{\text{course}_j}} u_{ij} + \sum_{j=1}^{n_{\text{class}_j}} v_{ij}) \tag{7}
\]

- Where \( crash_g \) is the number of conflicts of the population, \( reward_g \) represents the number of rewards and punishments of the population.
- Among \( u_{ij} \), represents the uniformity of teaching time distribution of the \( i \)-th course in the \( \lambda \) major, among them, \( n_{\text{major}} \) represents the number of specialties in Colleges and Universities, \( n_{\text{course}_j} \) is the total number of courses in major \( \lambda \).
- Where \( v_{ij} \), represents the uniformity of curriculum arrangement of the \( j \)-th class in the \( \lambda \) major, \( n_{\text{class}_j} \) indicates the total number of classes in major \( \lambda \).
- Where \( w_j \) is the priority of course time period arrangement, \( r_i \) represents the use rate of school classrooms?

### 7. Algorithm Operation Flow

According to the analysis of the scheduling problem, the specific operation process of the following algorithm is obtained.

Using the decimal coding method to code the schedule parameters and generate the schedule chromosome.

The number of initial populations is \( N \).

The fitness function is designed to cipher out individual fitness.

Execute selection operation and self-adapting variation and crossover operation.

Analyze whether the generated population can meet certain conditions. If the conditions are met, the best timetable will be output. If the conditions are not met, return to step 3.
8. Experimental Results of Improved Genetic Algorithm and Traditional Genetic Algorithm

According to the experimental data, 5000 students, 300 classrooms, 190 teachers, 550 courses and 100 classes were selected from the new campus of a university. The initial population 170, 3000 times of successive iterations had no change in fitness. Under the identical experimental conditions, the improved genetic algorithm in this paper and the traditional genetic algorithm are compared in terms of the evaluation data and the time consumed in the course arrangement.

8.1. Comparison Test Results of Running Consumption Time

The comparison of assessment criteria is an important link to verify the results of course arrangement. This paper gives the assessment criteria of classroom utilization rate, curriculum uniformity, curriculum leakage rate and curriculum conflict rate. According to the 3000 iterations of the experimental data, the improved genetic algorithm in this paper is contrasted with the traditional genetic algorithm in the assessment criteria.

| algorithm                | Classroom utilization | Course uniformity | Curriculum leakage rate | Curriculum conflict rate |
|--------------------------|-----------------------|-------------------|-------------------------|--------------------------|
| Traditional genetic algorithm | 75%                  | 0.63              | 5.32%                  | 6.46%                    |
| Improved genetic algorithm    | 89%                  | 0.78              | 1.54%                  | 2.17%                    |

Draw conclusions from the table that the improved genetic algorithm is superior to the traditional genetic algorithm in four aspects: classroom utilization rate, curriculum uniformity, curriculum leakage rate and curriculum conflict rate.

8.2. Comparison Test Results of Running Consumption Time

In accordance with 3000 iterations of empirical data, the time consuming of the improved genetic algorithm in this paper and the traditional genetic algorithm are contrasted. Draw a conclusion from the figure below, time consuming of the optimized genetic algorithm is less than that of the traditional genetic algorithm. Therefore, the improved genetic in this paper algorithm gets the globing first-best answer earlier than the traditional genetic algorithm.
9. Summary
This text primarily employs the decimal system of coding to code the chromosome of the timetable, this makes the coding method more minute and explicit. And the self-adapting crossover variation method is used to operate the gene, so it is easier to get the globing greatest result. The consequence reveal that the improved genetic algorithm in this paper is superior to the traditional genetic algorithm. Finally, the author hopes to be helpful to the scholars who study the scheduling problem.

Acknowledgments
The whole process of this paper from topic selection, data retrieval, writing modification to final finalization is full of doubts, pressure and fun. First of all, I should like to thank our respected tutor, Mr. Sun. Mr. Sun's real human nature, profound knowledge and rigorous scholarship spirit have benefited me a lot. It is under the care and guidance of Mr. Sun that I can successfully complete this project. I want to express my deep thanks to my tutor. In addition, I would like to thank my lab students for their encouragement and support. I hope this paper will be helpful to the scholars who study the scheduling problem, and I will be very pleased.

References
[1] Leonardo Cabral R. Soares, Marco Antonio M. Carvalho. Biased random-key genetic algorithm for scheduling identical parallel machines with tooling constraints[J]. European Journal of Operational Research, 2020.
[2] Xiong Luo, Qian Qian, Yun Fa Fu. Improved Genetic Algorithm for Solving Flexible Job Shop Scheduling Problem[J]. Procedia Computer Science, 2020, 166.
[3] Guohui Zhang, Yifan Hu, Jinghe Sun, Wenqiang Zhang. An improved genetic algorithm for the flexible job shop scheduling problem with multiple time constraints[J]. Swarm and Evolutionary Computation, 2020, 54.
[4] Jia Luo, Didier El Baz, Rui Xue, Jinglu Hu. Solving the dynamic energy aware job shop scheduling problem with the heterogeneous parallel genetic algorithm[J]. Future Generation Computer Systems, 2020, 108.