The application of Coarse-Grained Parallel Genetic Algorithm with Hadoop in University Intelligent Course-Timetabling System

Liping Wu
Shandong Agriculture and Engineering University, Jinan, Shandong, China

Abstract—**Course-timetabling problem is a NP-C problem at Universities. Many traditional methods for solving this problem show low-efficiencies in several aspects, such as a high conflict rate of teacher resources or classroom resources, low degree of satisfaction among students or teachers and so on. Thus these methods do not meet the requirements of modern university educational administration management. Parallel genetic algorithm (PGA) integrating the advantages of the traditional genetic algorithm (GA) and the computing power of parallel computing can improve effectively the quality and speed in solving problems. This paper, which is based on the cloud computing platform with Hadoop, presents an improved method fusing Coarse-grained parallel genetic algorithm (CGPGA) and Map/Reduce programming model, to solve the university course-timetabling problem. The simulation experiment results show that, compared with the traditional genetic algorithm, CGPGA not only improves the success rate but also reduces the conflict rate in solving the university course-timetabling problem. At the same time, this paper makes full use of the high parallelism of Map/Reduce to improve the efficiency of the algorithm. And then the corresponding computational results are reported.**

Index Terms—**Intelligent algorithm, Coarse-grained parallel genetic algorithm (CGPGA), Intelligent course-timetabling system, Map/Reduce**

I. INTRODUCTION

With the rapid development of economic and cultural level, the scale of university enrollment continues to expand, and the professional setting and course setting is also developing in depth and breadth. At the same time, the universities are faced with the scarcity of educational resources, especially in the classroom resources. Therefore, course-timetabling becomes one of the most difficult and time-consuming work of academic management at universities [1]. Course-timetabling problem is a dynamic combinatorial optimization problem involving many factors, such as teachers, classes, course, classrooms, times periods and so on, and it obviously had been proved to be NP complete (NP-C) problem [2]. Many traditional methods for solving this problem show low-efficiencies in several aspects, such as a high conflict rate of teacher resources or classroom resources, low degree of satisfaction among students or teachers and so on. So, these methods do not meet the requirements of modern university educational administration management. With the development of campus network and office automation, the university educational administrators realize that the construction of computer intelligent course-timetabling system is imperative [3-6].

Domestic and foreign scholars had carried out a large number of research works on the university intelligent course-timetabling problem. Many intelligent optimization algorithms had been successfully applied to the university intelligent timetabling problem because of its robustness and universality. Li Hongchan, Hu Gang and Zhu Haodong analyzed University Timetabling Problem detailed, established an optimization mathematical model, and designed a variety of improved schemes to solve UTP more effectively [7]. Tang Yong, Tang Xuefei and Wang Ling used genetic algorithm to optimize the initial course timetabling, and used matlab to programming. The test verified that the algorithm played a significant role in timetabling optimization [8]. YU Chengmin, JIANG Hua, LI Huan and JIA Baoxian presented an improved discrete Particle Swarm Optimization (PSO) algorithm to tackle the university course-timetabling problem, and verified the feasibility of the method [9]. According to the requirement and character of the course timetabling problem, Zhang Lin presented an algorithm based on the bipartite graph theory, and then he mixed the three different algorithms together and got an improved ant colony algorithm which can obviously improve the space of getting the answer and the quality of the answer [10]. David Abramson, Mohan Krishna Amoorthy and Henry Dang described the use of simulated annealing (SA) for solving the school timetabling problem and compared the performance of six different SA cooling schedules [11]. Ding Zhenguo and Zhao Hongwei introduced a method to solve the course-timetabling problem which mixed the classical network flows algorithm and the modern heuristic taboo search algorithm; The method fused the superiorities of two algorithms together and improved the ability of processing problems [12] .Wang Lu and Qiu Yuhui proposed the framework of intellective class scheduling system to research theories and technologies in Multi-Agent Systems; And the study aimed to concern about majority teachers' expectations and it intended to solve teacher' s satisfaction problem which is based on the negotiation techniques of Multi-Agent Systems [13]. Nie Xiaodong tried to use the greedy arithmetic to solve the problem to aim the running of course-timetabling system [14]. The study developed a system which is based on the greedy arithmetic, resource matching method and the dynamic memory distributing best adapting method. The response time and the result of the system obtained by this method had reached satisfaction. In summary, all kinds of intelligent optimization algorithms have been applied to solve the university intel-

ligent course-timetabling problem, and the most widely used is Genetic algorithm (GA).

With the expansion of the scale of the problem and the complexity of the search space, the traditional GA is prone to premature convergence, local optimal solution, low-efficiency and so on. In this paper, based on the cloud computing platform with Hadoop, an improved method fusing Coarse-grained parallel genetic algorithm (CGPGA) and Map/Reduce programming model is presented to solve the university course-timetabling problem.

The rest of the paper is organized as follows. Section 2 designs Coarse-grained parallel genetic algorithm which is based on Hadoop. Section 3 builds the mathematical model of intelligent course-timetabling system at universities. We carry out an experiment in Section 4 and its results analysis is given after the comparison of the improved genetic algorithm and the traditional genetic algorithm. The last two sections are the conclusion and the references.

II. DESIGN OF COARSE -GRAINED PARALLEL GENETIC ALGORITHM WITH HADOOP

Hadoop is an open source project which is developed by Apache software foundation. Users can set up a large scale cluster system by using it on the common hardware. Hadoop is an optional distributed system infrastructure. Its cores are HDFS (Hadoop Distributed File System) and Map/Reduce.

Coarse-grained model is also known as Distributed style or Island-based model, which is one of genetic algorithm parallel models with the most adaptive and widely used [15]. The results are improved by increasing the population diversity, which not by the calculation of the adaptation.

The resolving idea of Map/Reduce which is based on CGPGA with Hadoop is that the serial genetic algorithm is changed into a Map/Reduce operation. The calculation of the individual fitness value and the genetic operation of crossover and mutation is calculated in Stage Map; Stage Reduce is to determine whether to meet the convergence condition, if "yes" then the results are output, or we enter the next Map/Reduce operation. Unlike general Map/Reduce operation, the Map/Reduce of CGPGA is achieved by Stage Partition after Stage Map. The optimal individual is transferred by circular algorithm between the sub populations and the others remain independent.

In order to ensure that all the sub groups alone multiply, we define that the number of nodes in Stage Map and Stage Reduce is n, and the data of Stage Map are processed in corresponding Stage Reduce. Each individual await is to give a sub group key, in the Stage Map, the optimal individual's key is (key+1)mod n, and key mod n is operated in Stage Partition, so as to achieve the optimal individual circular migration. The process is shown in figure 1.

III. MATHEMATICAL MODEL

A. Model description

The problem of the course-timetabling at universities is mainly related to five factors: teachers, classes, courses, classrooms and times periods. N_p, N_c, N_l, N_r and N_t represent respectively the total number of all teachers, classes, courses, classrooms and times periods, the model encoding described as follows:
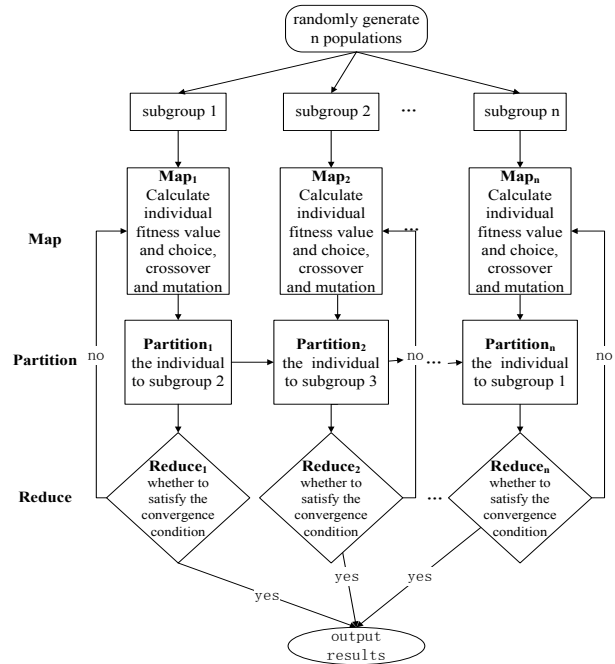

Figure 1. Design of coarse-grained parallel genetic algorithm with Hadoop

Set of all teachers is P ($P = \{P_i\}, i = 1,2,\ldots,N_p$) and the number of courses corresponding to each teacher is K ($K = \{K_{P_i}\}$).

Set of all classes is C ($C = \{C_{i'}\}, i' = 1,2,\ldots,N_c$) and the number of students corresponding to each class is G ($G = \{G_{ci}\}$).

Set of all courses is L ($L = \{L_{i'}\}, i'' = 1,2,\ldots,N_l$) and the number of classes corresponding to each courses is X ($X = \{X_{lr}\}$).

Set of all classrooms is R ($R = \{R_{i'''}\}, i''' = 1,2,\ldots,N_r$) and the number of desks corresponding to each classrooms is Y ($Y = \{Y_{ri}\}$).

Set of all times periods available is T ($T = \{T_{i''''}\}, i'''' = 1,2,\ldots,N_t$).

Cartesian product of times periods and classrooms is M (M = T·R = $\{(T_1,R_1), (T_2,R_2),\ldots,(T_{Nt},R_{Nr})\}$), then the course-timetabling problem is transformed into finding a suitable "time-classroom" for each class. There are total 2^{Nr} mapping relationships, which are as follows:

$$P_i C_i L_{i'} R_{i'} T_{i''} = \begin{cases} 1, \text{professor } P_i \text{ teaches course } L_{i'} \text{ for class } C_i \text{ in room } R_{i'} \text{ at time period} T_{i''} \\ 0, \text{otherwise} \end{cases} \quad (1)$$

B. Objective function

The course-timetabling problem can be considered as a resource allocation problem, the main task of which is that the classrooms, teachers, classes and courses do not appear to conflict within a week [16]. The resource allocation effects of the universities intelligent courses-timetabling are reflected in the following aspects:

- Teaching effect f_1: the important courses need to be arranged in good teaching unit.

- Teachers' needs to meet f_2: the teachers' requirements of the times periods, classrooms and so on must be meet.

- Course scheduling optimization in multi class hours f_3: the time interval of the course include more than multi class hours in a week should be as far as possible to above one day to ensure the teaching effect.
- Classroom resource utilization f_4: the number of classrooms' seats is equal to the number of students in classes.

In this paper, the goal of university intelligent course-timetabling problem is to achieve the maximum weighted sum of the resource allocation effect under all constraints, the hypothesis is as follows:

- The teaching unit of the university is divided into 5, α_i (i=1,2…,5) is defined as the good and bad effect of the teaching units. Where i=1,3,5, α_i =1, otherwise, α_i =0. The university courses are divided into elective courses, basic courses, professional courses and degree courses, β_j (j=1,2,3,4) is defined as the important degree of the courses, whose corresponding values are 1, 2, 3 and 4, then the optimization objective of the teaching effect is:

$$\max(f_1) = \sum_{i=1}^{5}\sum_{j=1}^{4} \alpha_i\beta_j \qquad (2)$$

- χ_i (i=1,2,3,4) is defined as teacher title coefficient whose values are 1, 2, 3 and 4, which corresponding to the title as assistant, lecturer, associate professor and professor. δ_j (j=1,2,3) is as the teacher willingness in each teaching units and classrooms, etc., the values of which are 0, 1 and 2, which corresponding to the willingness as unwilling, no matter and willing, then the optimization objectives of meeting the teachers' needs is:

$$\max(f_2) = \sum_{i=1}^{4}\sum_{j=1}^{3} \chi_i\delta_j \qquad (3)$$

- φ_i (i=1,2,3,4) is defined as the teaching effect coefficient of a course where the interval day is i days, whose values respectively are 1, 3, 4 and 2, which corresponding to the effect as poor, good, very good and general, then the optimization objective of the course arrangement is:

$$\max(f_3) = \sum_{i=1}^{4}\sum_{j=1}^{4} \varphi_i\beta_j \qquad (4)$$

- γ is defined as the number of the students learning the course and η is as the number of the seats in the classroom, then the optimization objectives of the classroom resource utilization rate is:

$$\max(f_4) = \sum {\gamma}/{\eta} \qquad (5)$$

The classroom resources utilization rate of each course is the bigger the better, and the maximum value is 1.

To sum up, the objective function is:

$$F = \sum_{i=1}^{4}\lambda_i f_i = \sum_{i=1}^{4}\lambda_i f_i \qquad (6)$$

In the formula, λ_i is as each target's weight, and we define to 3, 1, 2 and 4.

C. Constraint condition

The main constraints of the university intelligent course-timetabling problem are as follows:

- Each class is only arranged one course in one time period, that is:

$$\sum_{i=1}^{N_p}\sum_{i''=1}^{N_l}\sum_{i'''=1}^{N_r} P_iC_{i'}L_{i''}R_{i''}T_{i'''} \leq 1 \qquad (7)$$

- Each teacher can only teach one course in one time period, that is:

$$\sum_{i=1}^{N_c}\sum_{i''=1}^{N_l}\sum_{i'''=1}^{N_r} P_iC_{i'}L_{i''}R_{i''}T_{i'''} \leq 1 \qquad (8)$$

- Each classroom can only be arranged for one class, that is:

$$\sum_{i=1}^{N_p}\sum_{i'=1}^{N_c}\sum_{i''=1}^{N_l} P_iC_{i'}L_{i''}R_{i''}T_{i'''} \leq 1 \qquad (9)$$

- The number of students allocated to the classroom should be greater than or equal the number of students in the class, that is:

$$Y_{r_{i''}} \geq G_{c_{i'}} \qquad (10)$$

IV. EXPERIMENT AND RESULT ANALYSIS

A. Algorithm scheme of universities intelligent course-timetabling

According to mentioned theories above, the coarse grain genetic algorithm program of university intelligent course-timetabling is as follows:

- Use a binary for encoding and mapping "time-classroom".
- Uniformly generate an initial population, ensure that the initial solution is uniformly distributed in the solution space.
- The fitness function is the objective function.
- Selection operation

Roulette wheel method is be used. The higher the individual adaptation value, the greater the proportion of the roulette wheel and the probability of survival, that is the greater chance of entering the next generation.

- Adaptive crossover and mutation operation

The probability of crossover p_c and mutation p_m is calculated as follows:

$$p_c = \begin{cases} \dfrac{k_1(F_{max} - F_{avg})}{F_{max} - F'}1, & F' \leq F_{avg} \\ k_1, & F > F_{avg} \end{cases} \qquad (11)$$

$$p_m = \begin{cases} \dfrac{k_2(F_{max} - F_{avg})}{F_{max} - F}1, & F \leq F_{avg} \\ k_2, & F > F_{avg} \end{cases} \qquad (12)$$

In this formula, k_1 and k_2 are defined as the same random number (0, 1), F_{max} is the fitness value of the best individual in the current population, F_{avg} is the average

fitness value in the current population, F' is the large fitness values of individuals who participate in cross operation, and F is the large fitness value of individual in the crossover operation.

B. Simulation experiment

The data of the experiment is from a university course table, the realization method of the experiment is C# programming, and the distribution of various elements is shown in Table 1.

Experimental parameters of the coarse grain parallel genetic algorithm based on Hadoop are shown in Table II.

We use the improved algorithm to make a parallel experiment comparison with the traditional genetic algorithm.

The experiment is carried out 10 times, the best fitness value of the population is recorded one time of each 100 generations, and the experimental results of the fitness value is from the average value of the best fitness value in 10 times records.

Ten average values which are calculated by the method in the paper are better than that by the traditional genetic algorithm from Figure 2. It shows that using this method to solve the problem of universities intelligent course-timetabling get better effect.

The operating results by the two algorithms are shown in Table III.

By using this method, the overall conflict rate is reduced, the success rate is 100%, and the time consuming of searching for the optimal scheduling scheme is greatly reduced from Table III. In a word, the experimental results are satisfactory and this method is a better method in intelligent course-timetabling system.
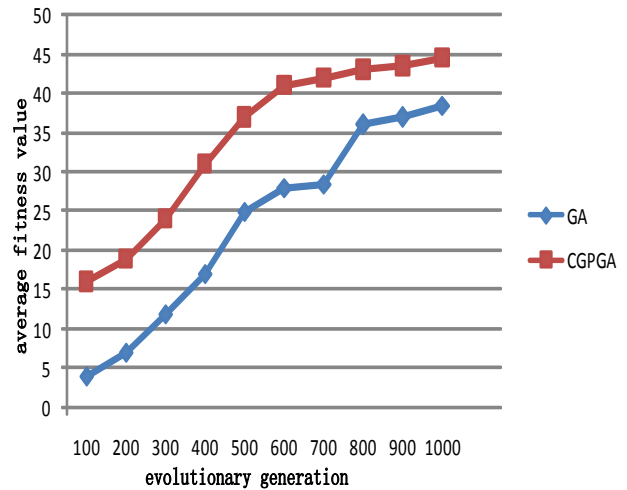

Figure 2. The fitness comparison

V. CONCLUSION

On the basis of analyzing the problem of the course-timetabling at universities, the mathematical model is constructed and the framework of the problem is given. In this paper, we use the coarse-grained parallel model which is based on Hadoop to mix the advantages of genetic algorithm to solve the problem of university intelligent course-timetabling. This method has reduced the workload of the educational administration personnel, and it also has a great effect on the teaching management in universities. Thus, it has extended value.

Although this method has some advantages in terms of computation results and operation efficiency, it introduces more parameters to increase the uncertainty of the algorithm. For example, how to determine the appropriate transfer mode becomes one of the problems to be solved in the future research.

TABLE I.
THE DISTRIBUTION OF VARIOUS ELEMENTS

elements	numbers
student	8000
professor	500
class	175
classroom	200
course	720

TABLE II.
THE EXPERIMENTAL PARAMETERS

parameters	numbers
initial population	200
subgroup n	4
maximum evolutionary generation	1000
selection probability	0.50
crossover probability	adaptive
mutation probability	adaptive

TABLE III.
THE ALGORITHM PERFORMANCE COMPARISON

performance indexes	CGPGA	GA
course conflict rate	0	30%
arrangement success rate	100%	86%
consume time	60 seconds	500 seconds

REFERENCES

[1] Adewumi A O, Sawyerr B A, Montaz A M. A heuristic solution to the university timetabling problem [J]. Engineering Computations, 2009, 26(8): 972-984. http://dx.doi.org/10.1108/02644400910996853

[2] S.Even. On the Complexity of Time-Table and Multi-commodity Flow Problems [J].SIAM Journal on Computing, 1976. 5(4): 691-703. http://dx.doi.org/10.1137/0205048

[3] Shi Juan. Research on application of IGA (Immune Genetic algorithm) to the solution of course-timetabling problem[C]//Proceedings of the 4th International Conference on Computer Science and Education, Nanning, China, 2009.

[4] Aladag C H, Hocaoglu G, Basaran M A. The effect of neighborhood structures on tabu search algorithm in solving course timetabling problem [J].Expert Systems with Applications, 2009, 36 (10):12349-12356. http://dx.doi.org/10.1016/j.eswa.2009.04.051

[5] Zhang D, Liu Y, Hallah R M.A simulated annealing with a new neighborhood structure based algorithm for high school timetabling problems[J].European Journal of Operational Research,2010,203(3):550-558. http://dx.doi.org/10.1016/j.ejor.2009.09.014

[6] Anmar A, Masr A. Multi-neighbourhood particle collision algorithm for solving course timetabling problems[C]//Proc of 2nd Conference on Data Mining and Optimization, Malaysia, Selangor, 2009: 21-27.

[7] LI Hongchan, HU Gang, ZHU Haodong. Research of UTP based on population dominant GA. Computer Engineering and Applications, 2011, 47(10): 233-236.

[8] Tang Yong, Tang Xuefei, Wang Ling. Timetable System Based on Genetic Algorithm [J]. Computer Applications, 2002, (10):93-97.

[9] YU Chengmin, JIANG Hua, LI Huan, JIA Baoxian. Research of Course Schedule Problem Basedon Discrete Particle Swarm Optimization Algorithm [J]. Microcomputer Applications. 2010, 31(7):67-71.

[10] Zhang Lin. Research and design of courses arrangement based on ant colony algorithm [D] .Hefei: Anhui University, 2005.

[11] David Abramson, Mohan Krishna Amoorthy, Henry Dang, Simulated annealing cooling schedules for the school timetabling problem [J], Asia- Pacific Journal of Operational Research, 1999(5).

[12] DING Zhenguo, ZHAO Hongwei. Research on Application of TS Algorithm to the Solution of Course-Timetabling Problem [J].Microelectronics & Computer, 2008, 25(4): 31 -34.

[13] Wang Lu, Qiu Yuhui. Research of Intellective Class Scheduling System Based on Negotiation [J]. Computer Science, 2006(6): 214 -217.

[14] Nie Xiaodong. The Research and Implementation of arranging course system based on greedy arithmetic [D]. Guangdong: Guangdong University of Technology, 2006.

[15] Corcoran A L, Wainwright R L. A Parallel Island Model Genetic Algorithm for the Multiprocessor Scheduling Problem[C]//Proceedings of the 1994 ACM/SIGAPP Symposium on Applied Computing. USA: ACM Press, 1994: 483-487.

[16] Pillay N, Banzhaf W.A study of heuristic combinations for hyperheuristic systems for the uncapacitated examination timetabling problem [J].European Journal of Operational Research, 2009, 197(2): 482-491. http://dx.doi.org/10.1016/j.ejor.2008.07.023

AUTHORS

Liping Wu is a lecturer at School Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan, Shandong 250100China. Her current research interests include Computer software and theory. (e-mail: wlp5@163.com).