2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE 2021)

Research on Software Project Scheduling Based on Genetic Algorithm

SiChong Ma[1] Shi Song Deng[2]
College of Management, Hefei University of Technology, Hefei, Anhui 230009

Absrtact: **software project scheduling problem (PSP) is a resource scheduling problem that arranges the optimal human resources for the project and satisfies the project constraints under the constraints of task priority and skill constraints. Because the core resource of software project scheduling problem is human, there are skill constraints, maximum contribution value constraints, and priority scheduling of tasks. In this paper, genetic algorithm is used to abstract a certain number of candidate solutions by binary coding, and the population is advanced to a better solution by iteration. By using TPG graph to represent the priority between tasks, the crossover operator and mutation operator are used to avoid falling into the local optimal solution, and the standardized method is used to prevent the employees from working overtime. The fitness function is used to select the optimal solution in the population. Experiments show that genetic algorithm has good applicability and effectiveness for project change and fitness function change.**

Keywords: software project scheduling problem, genetic algorithm, multi-objective, normalization

I. INTRODUCTION

In our real life, there are always some medium or large-scale projects, in which project scheduling is our concern, how to allocate human resources effectively, and get a satisfactory solution of time and cost. A project scheduling problem, we often have to involve different tasks, employees with different skills, when the number of tasks and skills, the project manager is difficult to allocate personnel effectively, resulting in a waste of resources. It is very difficult to do this alone. We need to study computer-aided tools to plan the rational [2][3] of project development.

In this paper, we mainly use genetic algorithm to solve the problems in project scheduling. The structure of the article is as follows. In chapter 2, we model the problem. Chapter 4 discusses the improvement of the algorithm, chapter 5 sets the parameters and chapter 6 summarizes and prospects.

II. PROBLEM MODELING

The software project scheduling problem (PSP) currently refers to the resource-constrained project scheduling problem (RCPSP), which is the problem of finding the optimal plan for the software project under the condition of satisfying the project priority and resource constraints. The problem can be described as:

The problem of software project scheduling is to assign employees with different skills to j tasks of the project and to meet the skill matching relationship between the employee and the task (the skill set of the employee matching to the skill set required by the task). Reduce project cost and completion time while meeting constraints.

The mathematical models used in this paper are as follows:

The objective function is:

$$min\, Q = w_{cost} \cdot cost + w_{time} \cdot time$$

The constraints are:

$$req_j \subseteq \bigcup_{i=1}^{n} \{skill_i \mid x_{ij} > 0\} \quad (1)$$

$$d_{i,j} = \frac{x_{i,j}}{max(1, \Sigma_{d_t \in v'} x_{i,t})} \quad (2)$$

$t_j e_i$ Among them, formula (1) indicates that the skill set required by the task must be a union of all employee skill sets; formula (2) indicates that employees are not allowed to work overtime, and once overtime behavior occurs, it must be standardized.

Because there are certain constraints between each task in the project, these constraints indirectly specify the order between the tasks, and all the constraints and order will be reflected in the task priority chart (TPG.As a result, all constraint relationships can be defined in the task-first graph (TPG)(shown in Figure 1).

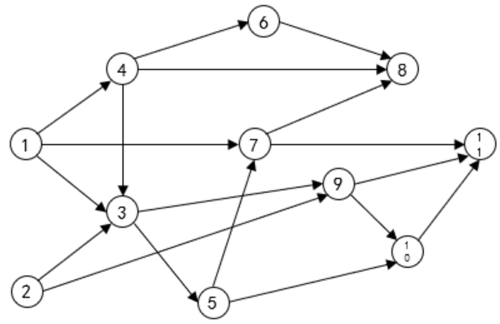

Figure 1 TPG Figure

III. GENETIC ALGORITHMS

A. Optimal framework for genetic algorithms

1) improve the structure of genetic algorithm

Improvement of selection operations

In order not to destroy the optimal solution of contemporary individuals because of the randomness of cross variation, the individuals with the highest contemporary fitness can be directly retained to the next generation in the selection operation. Or in order to ensure that the individuals whose fitness is higher than the average can get a certain reservation, the combination selection mode of optimal choice and competitive choice is put forward.

Improvement of cross-selection

2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) | 978-1-6654-1540-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICBAIE52039.2021.9389854

Try to make the final can wait until the satisfactory solution, in the intersection should pay attention to. The introduction of crossover can often change the convergence rate and solution diversity of the population, but with the increase of the population number, in order to retain the solution with large fitness, we can use the cross introduction of smaller points. In addition, when crossing, it may be considered to avoid inbreeding, or to use the crossover operator of clustering.

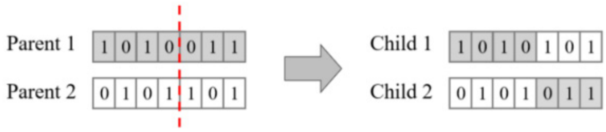

Figure 2 Cross

Improvement of mutation operation

In order to ensure the diversity of the results, the probability of variation becomes particularly important. the probability of our mutation will decrease with the increase of algebra. since at first, the fitness of the population is not high, a large probability of variation can guarantee us to get more solutions. as the algorithm progresses, this probability will tend to a stable value and try not to destroy the solution [4] with high fitness. For example, the following mutation probability:

$$p_m m_2 \times m_1 = \cos(+t) + P_{ml} \quad (3)$$

where the t is the number of iterations.

2) improve the problem model of genetic algorithm Introduction of overtime workload settings

In a real life, overtime often occurs, but excessive overtime may lead to a decline in the correct rate of employees, from making task completion time longer. At the same time, appropriate overtime can be allowed to estimate the time and cost of project completion more closely. Of course, when overtime exceeds a certain limit, it will give a larger penalty value.

$$(x)\begin{cases} \frac{1}{q}, & \text{meet the question} \\ \frac{1}{q+p}, & \text{else} \end{cases} \quad (4)$$

$q = w_{cost} \times cost + w_{time} \times time;$

$p = w_{penal} + w_{undt} \times undt + w_{reqsk} \times reqsk + w_{over} \times over.$

IV. Improvement of Algorithm

A. Standardized working hours

In the previous literature, genetic algorithm was directly used to solve the optimal solution, but the problem of employee overtime was ignored. In real life, employees working overtime is generally not considered. In order to solve this problem, we propose a standardized method. That is, once the employee overtime work occurs in the solution (the contribution value of the current executable task is greater than 1), then the contribution degree is normalized by the formula to equal to 1.$d_i d_i$

$$d_{i,j} = \frac{x_{i,j}}{\max(1, \Sigma_{d_t \in v'} x_{i,t})} \quad (6)$$

The use of standardized methods can be used to eliminate the key obstacles that affect the search process of the algorithm and alleviate the excessive work of software personnel. Because employees can distract attention from multiple jobs. Standardization also applies to smaller-grained schedules, because employees can automatically redistribute their contributions when a job is done or a new job starts. On the other hand, employees in a single job, will cause waste of resources. after normalization, the significance of the contribution value changed slightly. For any employee who does two or more at the same time, the same contribution means that he needs to divide his energy equally into all work. Through standardization, although the contribution value has changed slightly, for employees who need to contribute to two or more tasks at the same time, The standardized contribution means that he will allocate his energy reasonably on each task and ensure that he does not work overtime while fulfilling his performance.

B. Improvement of Selection Operator

Choosing a binary tournament will inevitably screen out individuals with greater fitness than the average. In order to ensure that these individuals can be retained to the next generation, the population is sorted according to the fitness value ,1/4 of the poor fitness individuals are eliminated directly, and 2/4 of the moderate fitness individuals are directly retained as the next generation.

Algorithm 1Select

Evaluate(solution)

Output: (new_solutions)

```
Let n =solutions.length
for solution in solutions do
    sort solution by fitness value from largest to smallest
        for i<=n/4-1do
            for j<=i do
            new_solutions j =solution i
            end for
          for n/4-1<j <=n/2-1do
              new_solutions j =solution i
        end for
      end for
    n/4-1do for n/4<i <=3
      for n/2-1<j <=n-1do
        new_solutions j =solution i
      for 3n/4-1<i <=n-1do
    delete solution i
```

The output result of algorithm 1 will eliminate 1/4 individuals with poor fitness and completely retain all the individuals in the top 3/4 of fitness as the next generation. In theory, this treatment can increase the probability of excellent individuals multiplying to produce better individuals.

V. EXPERIMENTAL RESEARCH AND ANALYSIS OF RESULTS

A. *Parameter settings*

Table 1 Parameter Initialization Settings

Parameters	Definition	Value
generation	Number of iterations	2000
selectNumber	Selector Selection Number	32
crossoverRate	Cross rate	0.6
mutateRate	Variance rate	0.3
piece	Time segmentation units	7
	Weight of costs	1
	Weight of time	1

Here, several optimal solutions are given so that readers can better understand the characteristics of the algorithm, as shown in the following table:

Table 2 Initializing the optimal solution

Num	Cost	Time
1	467758.81	19.6666
2	471357.03	19.5454
3	471588.53	20.3333
4	473055.35	30.3333
5	476588.21	19.6666
6	467990.31	20.3333
7	477885.40	20.3333
8	469747.01	18.0000
9	479557.28	30.2222
10	471588.53	20.3333

In initialization settings, we have the same weight for time and cost settings, so the structure of the optimal solution fluctuates more in a certain region and tends to stabilize. The fluctuation of the optimal solution depends on the crossover operator and mutation operator in the iterative process.

Next, we will test the ability of genetic algorithm to find the optimal solution by adjusting different parameter values, and compare the advantages of different parameter combination solutions.

B. *Changes in crossover rates*

Table 3 Optimal solution with a crossover rate of 0.3

Num	Cost	Time
1	482888.57	44.0000
2	474098.10	27.0000
3	457589.82	55.6666
4	479290.36	34.0000
5	472507.54	49.3333
6	463261.05	24.6666
7	463176.23	34.0000
8	457778.91	34.0000
9	485459.13	43.3333
10	446983.74	39.0000

Table 4 Optimal solution with a crossover rate of 0.8

Num	Cost	Time
1	472989.99	19.6666
2	469747.01	18.0000
3	469747.01	18.0000
4	472823.85	29.3333
5	473345.22	18.0000
6	475854.79	24.6666
7	473345.22	18.0000
8	474445.34	25.5000
9	465684.31	20.3333
10	485554.46	21.0000

When the crossover rate is high, the optimization rate increases and is close to the global optimal solution, but in the running process, the running time is longer and the resource is occupied.

C. *mutation rate changes*

Table 5 Optimal solution with a mutation rate of 0.1

Num	Cost	Time
1	478649.25	29.8666
2	473280.28	66.6944
3	466167.42	60.9777
4	469934.34	48.5333
5	471069.40	39.7500
6	472243.61	65.3333
7	467758.81	19.6666
8	460805.69	80.8888
9	472243.61	65.3333
10	473770.30	36.5000

Table 6 Optimal solution with a mutation rate of 0.5

Num	Cost	Time
1	477885.40	20.3333
2	474812.05	28.0000
3	474812.05	28.0000
4	473345.22	18.0000
5	473345.22	18.0000
6	473345.22	18.0000
7	475897.20	22.0000
8	480668.87	32.0000
9	471357.03	19.6666
10	480209.37	28.0000

A lower mutation rate may cause the population to fall into the local optimal solution, thus showing that the current optimal solution fluctuates greatly. On the contrary, a higher mutation rate is more conducive to finding solutions close to the global optimal solution or even the global optimal solution.

VI. SUMMARY AND OUTLOOK

Genetic algorithm is a method to search the optimal solution by simulating the natural evolution process, which has been well applied in the software project scheduling problem. Based on the previous literature, this paper puts forward some improvement schemes of the software project scheduling problem based on genetic algorithm. In addition, it also puts forward the standardized method of working hours for employees' overtime, which is closer to real life. In the simulation experiment, we adjust the relevant parameters by comparing the solution, and give the feasible optimization scheme for the project.

Genetic algorithm has good maneuverability, has been widely used in optimizing all kinds of problems. In this paper, genetic algorithm is implemented to solve the problem of software project scheduling, but there are still many aspects to improve. We put forward the following prospects:

(1) Presentation of new methods or strategies. In the future, other methods can be used to improve the algorithm itself, such as selecting individuals through roulette algorithm, so that individuals with better fitness can be retained to the next generation.

(2) Optimizing the problem using other intelligent algorithms. In the future, we can consider combining genetic algorithm with particle swarm optimization algorithm, which may get faster convergence speed and improve local search ability.

(3) Consider more complex issues. This article does not consider too many dynamic factors, such as employee turnover, the arrival of emergency tasks and so on.

ACKNOWLEDGMENT

Fund project: innovative entrepreneurship training program for college students (No.S201910359161)

Author

Si Chong Ma, （2000-), undergraduate, intelligent algorithm in main research field,1280888758@qq.com

Shi Song Deng,（2000-), undergraduate, intelligent algorithm in main research field,1294862909

REFERENCES

[1] Holland J H.Adaptation in natural and artificial systems.Univ of Michigan Press,Ann Arbor Mich,1975

[2] Bagley J D.The Behavior of Adaptive System which Employ Genetic and Correlation Algorithm[J]. Dissertation Abstracts International,1967,28(12)

[3] Yongzai Lu.Meeting the challenge of intelligent system technologies in the iron and steel industy.Iron and Steel,1996,(9)：139149

[4] Luo L F,Tsai L,Zhou Y M.Informational parameters of nucleic acid and molecular evolution[J]. J Theor Biol,1988,130：351361

[5] Fan Zenghui, Fang Wei. A Study of Improved Genetic Algorithms in Software Project Scheduling [J].]; and Computer Applications and Software ,2016,33(12):34-39+48.

[6] Li Xuehui. Optimization and Implementation of Project Scheduling Based D]. Genetic Algorithm Dalian University of Technology ,2015.

[7] Kong Lingfei. Application of Genetic Algorithms in Software Project Scheduling [D].]; and Jilin University ,2008.

[8] Zhang Qilin, Fan Liang. A Study on Resource Scheduling in Software Project Management [J].] in China Information and communications ,2013(10):25.

[9]Ke Li,Jian Wen,Yuce Liu,et al.Application of entransy theory on structure optimization of serrated fin in plate-fin heat exchanger.2020,173

[10]Saeed Rayegan,Shahrooz Motaghian,Ghassem Heidarinejad,et al.Dynamic simulation and multi-objective optimization of a solar-assisted desiccant cooling system integrated with ground source renewable energy.2020,173

[11]Han Wei,Hua Bao,Xiulin Ruan.Genetic algorithm-driven discovery of unexpected thermal conductivity enhancement by disorder.2020,71

[12]Jiayu Wu,Huiting Chen,Huiying Wang,et al.Will the opening community policy improve the equity of green