Modeling Resource-constrained Project Scheduling Problem and its Solution by Genetic Algorithm

Chunlai Chai
College of Computer Science and Information Engineering
Zhejiang Gongshang University
Postal code: 310018
P. R. China
ccl@mail.zjgsu.edu.cn

Journal of Digital
Information Management

ABSTRACT: *The optimization of the resource-constrained project scheduling is an NP-hard problem. Complexity of the algorithm for solving this problem increases exponentially with the increase of resource constraints. Therefore, traditional optimization methods based on Excel tables cannot calculate the efficient allocation of resources, which will affect the scheduling of project management resources. The optimization problem like multiple resource-constrained project scheduling demands an optimization method based on genetic algorithm, which adopts the decimal encoding based on activity priority combined with the storage of the adjacency matrix in order to effectively solve unauthorized phenomenon in activity scheduling. It also adopts preemption mode of allocation of resources in activities to avoid conflict in resource allocation. Besides, we designs C language program that realizes heuristic genetic algorithm, which can find out the most excellent duration that meets the resource constraints through multiple iterations. Finally, the practical results indicate that genetic algorithm can quickly and efficiently solve this problem.*

Categories and Subject Descriptors
K.6.1 Project and People Management I.1.2 Algorithms F.4.3 [Formal Languages] Classes by resources

General Terms:
Project Management, NP-hard problem, Genetic Algorithm

Keywords: Project Scheduling, Resource Constraints, Genetic Algorithm, NP-hard Problem

Received: 12 November 2012, Revised 28 December 2011, Accepted 31 December 2012

1. Introduction

Project is a special collection of a finite number of tasks within a certain time frame, and a general term for some related work in line with a series of specific requirements. Project management is a new approach of management by creating a temporary flexible organization, which would systematically manage and control the project and by high-efficiency project planning, organization, guiding and control would achieve a dynamic management throughout the whole process of project and a general adjustment and optimization of the project goals. With the rapid socio-economic development, the importance of project management has become increasingly prominent in production, management and other socio-economic activities. Especially with the emergence of new areas of project management such as network engineering, information system engineering, software engineering, development of high-tech projects and large construction projects since 1980s and 1990s, methodologies and theories about project management have been greatly and continuously developed.

Project scheduling problem is one of the core contents in project management, and is a process that allocates the existing resources reasonably to the scheduled activities, so as to achieve the intended goals under certain constraints (Usually the goal is to minimize the completion duration of the project). The research may be divided into several main stages: the stage of the Gantt chart, deterministic network planning stage, the critical path method (CPM) stage, probabilistic stage of planned network technology, such as PERT (program evaluation and review technique), the stage of optimizing the network constrained by multiple resources (such as human and financial resources, etc.), such as optimizing by adopting heuristic algorithms and optimization algorithms[1]. In China, Gantt charts, CPM and the PERT are adopted rather more widely, but those methods are mainly designed for single project management without any resource limit. However, a real project management process usually involves managing multiple projects simultaneously and the projects would affect each other and compete for resources. No consideration of the resource constraints means a schedule that is unattainable for lack of resources[2]. Therefore, how to coordinate the plan and resources of each project in case of multi-resource constraints has become a hot topic in current project scheduling study[3]. A Review about resource constrained project scheduling problem (RCPSP) was first put forward in 1963 by Kelley[4] when attempting to solve resource-constrained project scheduling problem. The basic resource-constrained project scheduling problem is a typical NP-hard problem. Mohring et al.[5] have pointed out that the RCPSP is one of the most difficult problems in the field of operation research. Many of the latest optimization technology and local search technology can be used to solve those kinds of problems, but due to its complexity and difficulty, the RCPSP has always been the focus of the project management scholars, and currently they mainly focus their research on the optimization algorithm. For example, Mao et al.[6] put forward an exact solution by adopting branch and bound algorithm that takes advantage of the event-driven time incremental method to solve the resource-constrained project scheduling problem; Shou[7] proposed using iterative algorithm to optimize the project scheduling problem; Cheng and Wu[8] worked out a hybrid intelligent algorithm by combining branch and bound algorithm and heuristic algorithm as a solution to such problems; Liu et al.[9] designed a buffer area on the non-critical chain in the project planning to optimize scheduling. Those algorithms all aim at the exact solution as the final optimization. Alcaraz and Maroto[10] think that those optimization algorithms can only yield an exact solution in small-scale projects and the number of its activities is usually small without a lot of resource constraints. At present, the optimization objectives in sense of resources/duration are mainly divided into two types: limited resources with the shortest period and balanced resources with a fixed duration. As to the optimization of limited resources with the shortest period, both the minimum duration and well balanced resource requirements can be achieved.

This paper intends to analyze the research status of the resource-constrained project scheduling problem. First it will analyze in case of resource constraints the problems in the project scheduling problem and study possible related mathematical modeling; second it will introduce genetic algorithms to solve resource-constrained project scheduling problems; in addition, this paper will try to improve the genetic algorithm by solving some defects in practical process, and attempt to design an algorithm structure with consideration of the actual characteristics of the RCPSP problem. Finally, a benchmark for project scheduling problem will be given in order to verify the effectiveness of the proposed model and genetic algorithm.

2. Modeling project scheduling problem

To establish the shortest project duration model in case of multi-resource constraints, the project is assumed to meet the following five assumptions:

(1) The number of activities is definite.

(2) Only a non-pre-empt mode of execution is allowed in activities.

(3) The constraint relationship of activities is determined to be an end - start constraint relationship.

(4) When activities require a variety of renewable resources, supply and demand amount of each kind of resources is determined.

(5) The implementation of activities is limited only by the constraints of renewable resources.

Based on the listed assumptions above, the mathematical model for project scheduling under resource constraints is as follows:

$$\text{Min } t_{n+1} \tag{1}$$

$$s.\ t.\ t >= 0 \tag{2}$$

$$T_j \geq T_i + d_i\ (j \in A_i) \tag{3}$$

$$\sum_{i=1}^{n} r_{ik} \leq R_k,\ i = 1, 2, ..., n,\ k = 1, 2, ..., m \tag{4}$$

In the formula, n represents the number of activities, activity 0 and activity $n+1$ are virtual activity, virtual activity 0 represents the start of activities, virtual activity $n+0$ represents the end of activities, t is the time of virtual activity, T is the time of actual activities, K represents the types of resources, t_{n+1} represents the starting time of virtual activity $n+1$, t_0 represents the starting time of virtual activity 0, d_i is the duration of activity i, T_i is the starting time of predecessor activity of T_j, r_{ik} is the daily demand amount of active i of resource k, R_k is the available amount of resources k, A_i is the collection of all predecessor activities of activity i. In the above model, equation (1) shows that the goal of scheduling is to minimize duration of the project, equation (2) defines that the project duration is positive, equation (3) indicates the preference constraints among activities, equation (4) shows the resource constraints that the activities are required to meet.

3. Genetic algorithm for solving this problem

The genetic algorithm provides a general framework for solving the optimization problems in complex systems. It does not depend on the specific areas of the problem. This paper applies genetic algorithm to optimize project scheduling problem, and by using heuristic group random search method in the search process it can avoid being trapped into local optimal situation, and thus can quickly and efficiently lead to the optimization scheme of the RCPSP. Evolution algebra is assumed to be G, and the genetic algorithm process is shown in Figure 1. Because the genetic algorithm is an off-line optimization algorithm, the conversion from the search space to the genetic space must be completed before the optimization, and there are mainly two steps of efforts. Step one is to convert the objective function into the fitness function. The RCPSP requires that the optimization goal be the minimum duration, which is a problem of a minimum number, while in the genetic algorithm the individual item that can maintain good viability should maintain the best fitness, which is a problem of a maximum number, so a conversion through ration computing is required. Step two applies chromosome coding to convert actual AoN (active on node) network into scheduling of project activities through random-based priority.

relationship, and at the same time reflects the random execution of the activities, and finally generates the coding of activity scheduling. The process is shown in Figure 2.

It uses the decimal encoding, and this kind of corresponding decoding is relatively simple. The so-called fitness value is calculated first by adopting the chromosome's corresponding scheduling and second by using the resources' preemption mode. Therefore, as long as the comparison between the best fitness values is made, the scheduling of the corresponding best chromosome can be searched, which is called the optimal

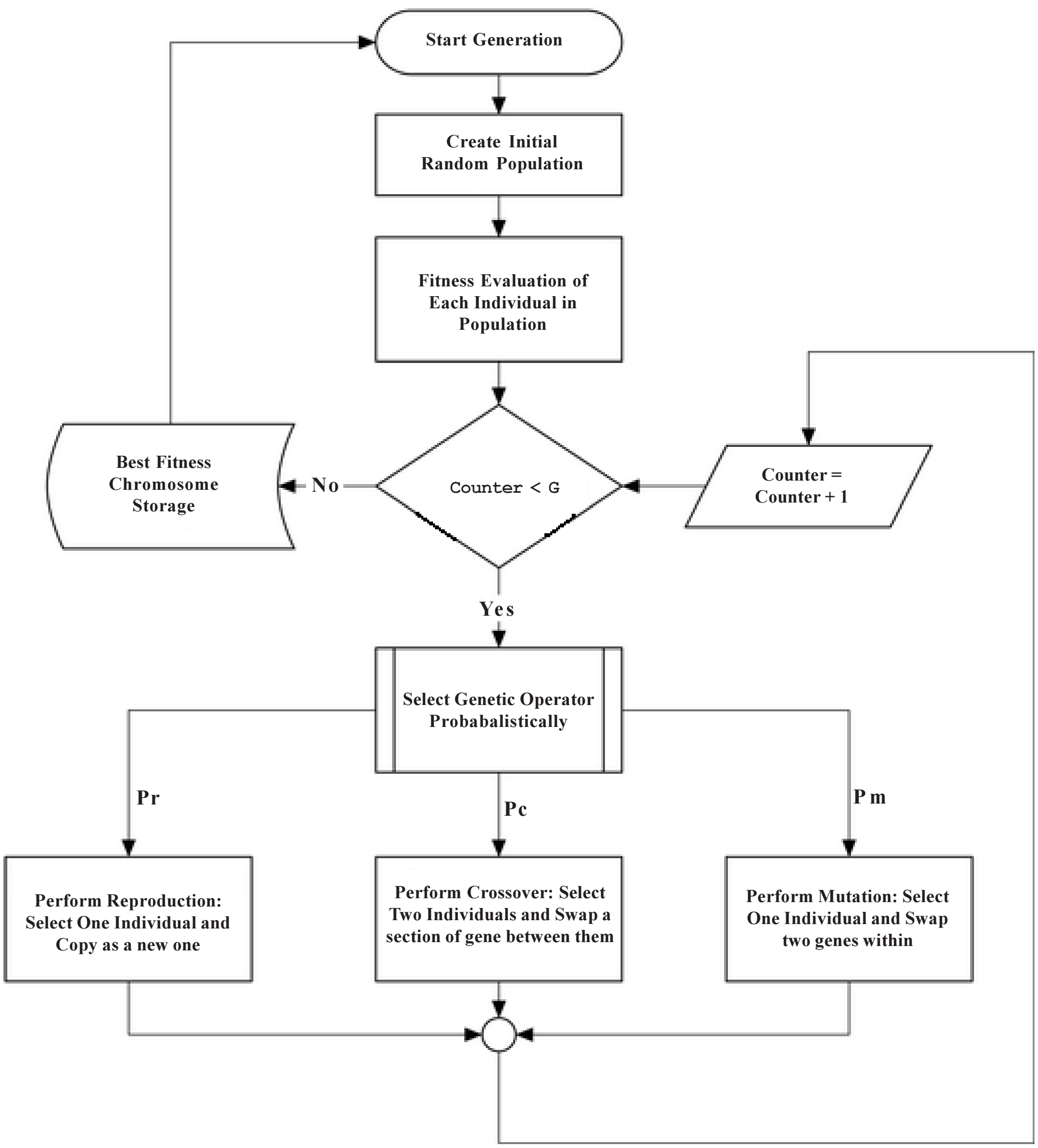

Figure 1. Genetic algorithm flow chart

3.1 Design of encoding and decoding

To apply genetic algorithm, the first problem to solve is encoding. It is a very critical step in the design of genetic algorithm, as the purpose of encoding is to make the scheduling of activities. This paper adopts an encoding method that is based on the sequence of priority activities. Its process follows the order of the activity constraint scheduling. The important process of converting the AoN network to the abstract coding space actually completes the generation of scheduling, which is done by using random priority code to search the AoN network. It can avoid code violation, and consists of three steps.

(1) The AoN network is indicated on computer, but it is

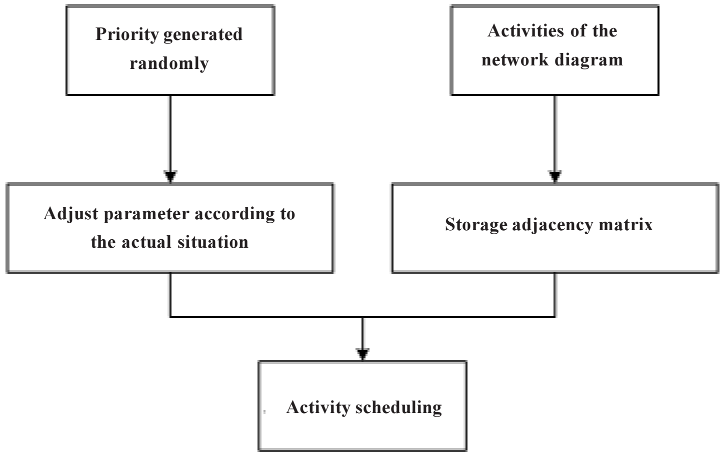

Figure 2. Activity scheduling encoding flow chart

achieved through storing adjacency matrix in the computer. The AoN network is a constraint basis for generating activity scheduling.

(2) The chromosome coding is generated by adopting the continuous random number of the priority activities, as shown in Table 1.

Number of Activities	1	2	3	…	n
Priority	P_{r0}	P_{r1}	P_{r2}	…	P_{rj}

Table 1. Sequence of chromosome

Each event has one and only one priority. As the priority that is generated randomly and that is determined by actual project activities may be different, the generated priority needs modifying.

The resulting priority sequence is actually a chromosome sequence, and each allele represents the priority of the corresponding activities. The chromosome code randomly generated by the computer must be $\{P_{ri} \in random\ (1, n+1)\}$ (P_{ri} *represents priority*), and in order to ensure the uniqueness of the project activity priority, the priority must not be repeated.

(3) The generation of activity scheduling
Scheduling of project activities is generated through the adjacency matrix of the activity storage and the activity priority. The fitness value depends on that step to be calculated, and the specific method is shown in Figure 3:

In addition, in order to ensure the effectiveness of successor node, the eligibility of each node must be determined. The judgment of eligible activities is based on two conditions:

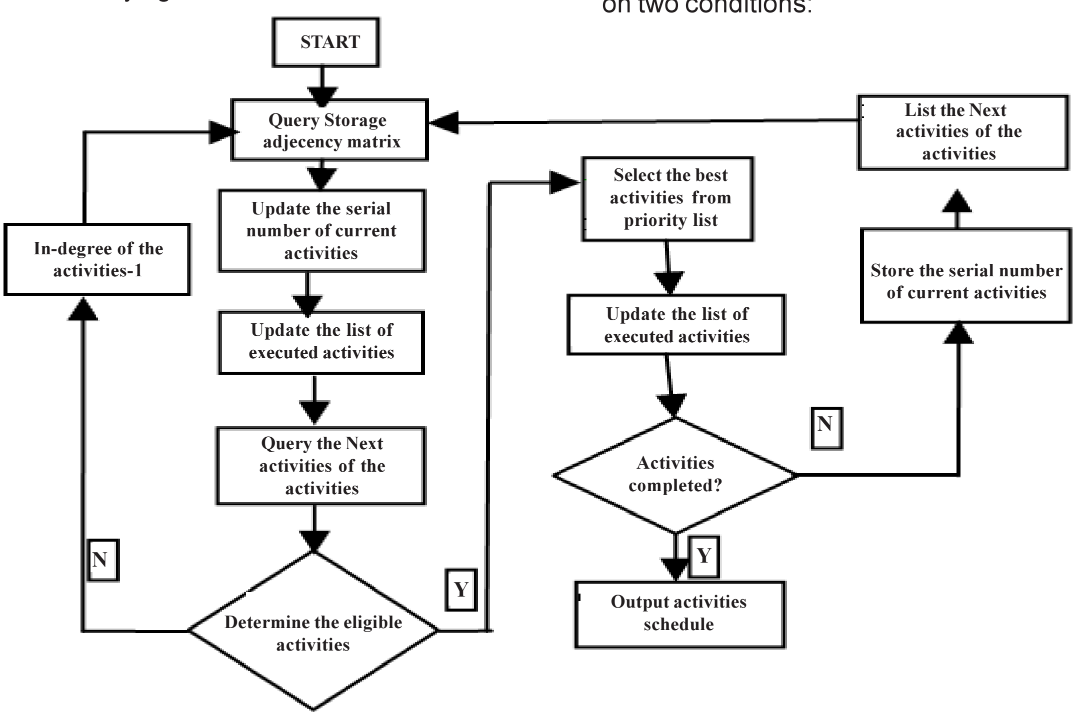

Figure 3. The generation process of activity scheduling

(1) Its predecessor activities have been implemented (the initialization of virtual activity 0 stands for implemented activities).

(2) The in-degree of the activities is computed according to formula (5), namely:

$$\lambda_{ID_j} = \sum_{i=j=1}^{n} C_{[c_i][c_j]} \quad (5)$$

In the formula, ID_j is the activity number, C is the activity's storage adjacency matrix, c_i is the line element, and c_j is the column element. As to those activities that meet all conditions, scheduling will be made according to their priority.

3.2 Design of fitness function

The fitness function is also called the evaluation function. It is determined by the objective function and serves as the criteria to distinguish between good or bad individuals in a group. It is also the driving force of the algorithm in the evolution and the sole basis for natural selection. Under the condition of resource constraints, the process to generate the target value and convert it into a fitness value is shown in Figure 4.

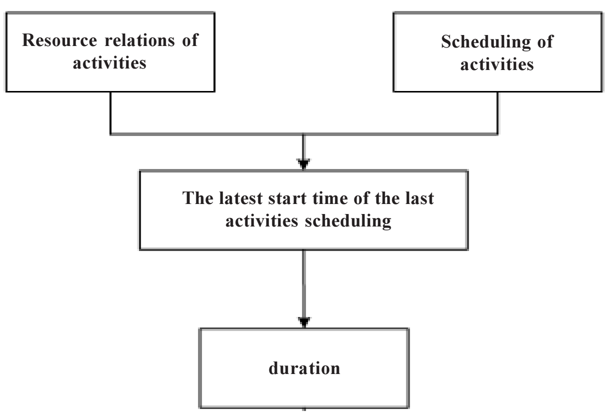

Figure 4. Calculation process of fitness value

3.2.1 The calculation of target value

The optimization goal of RCPSP problem is to find the minimum duration, that is also the latest start time of the last virtual activity. Because different scheduling would produce difference duration within the resource constraints, the calculation of the duration is mainly made according to scheduling and the resource constraints. The specific calculation steps are as follows.

Step 1: Initialize the resource list. The number of the initialized scheduling is equal to 1.

Step 2: Extract the activity ID number from the dispatch table, and find all relative predecessor activities. Select the ending time of the latest activity as the start time of the activity of that ID.

Step 3: Determine whether all relative resources meet the requirements. If yes then jump to Step 5.

Step 4: If no, then delay the earliest start time of the activity by one day, and then recalculate the supply of resources. If still no, then repeat steps 4.

Step 5: If yes, then start the activity. Make calculation according to the duration of the activity, allocate resources according to the preemption mode, delete all used resources from the resource list, and then update the list of resources.

Step 6: Determine whether all the activities have been allocated resources. If no, then add 1 to the serial number of scheduling, and return to step 2.

Step 7: The latest start time of the last activity (finished virtual activity) will be yielded, namely the target value or the duration.

3.2.2 The fitness function

As what is required is the shortest duration of activity, which is a minimum number, the original target value should be converted to a fitness value when designing the fitness function, in order to ensure that outstanding individual has a large fitness value. The scale transformation in equation (6) can convert the target value to fitness value, namely:

$$fitness(I) = \frac{D_{\max} - D_j + \gamma}{D_{\max} - D_{\min} + \gamma} \quad (6)$$

In the equation, I is the i-th chromosome in the current population, fitness (I) is the fitness value function, D_{max} is the maximum target value of the current population, D_{min} is the minimum target value of the current population, D_i is the target value to be converted, and γ is a positive real number within the open interval $(0,1)$. The introduction of γ can prevent the above equation from being divided exactly, and also allows adjusting selection behavior from the proportion of the fitness value to the pure random number selection. If the fitness value gap between the chromosome s is relatively large, then select according to the proportion of the fitness value; if the gap is relatively small, then select randomly among the potentially competing chromosomes.

3.3 Design of selection operation

Select in a roulette way according to the calculated fitness value from mentioned steps above. First generate a random number α between the interval [0, 1], then select according to equation (7), namely:

$$\sum_{j=1}^{n} \frac{f(x^j)}{\sum_{j=1}^{popsize} f(x^j)} \quad (7)$$

In the equation, n is the number of activities, and *popsize* is the group size. This equation will work out breeding parent group.

3.4 Design of crossover operation

The crossover operation simulates the biological process of natural evolution, where creatures match two homologous chromosomes, reorganize and form a new one, and thus produce new individuals or species. This study applies a two cross-computing model. The traditional two-point cross to cross model may produce weight code when crossing the locus of activity scheduling sequence, but this paper presents the two cross-location-based mapping. The first step in the crossover operation is the same with the traditional two-point cross, i.e. to exchange the corresponding allele in the 3th and 4th locus of parent 1 and parent 2; the second step is to establish a mapping relationship between the alleles in the corresponding locus of the two parents in the first step. Operate in the same way with other cross points and will finally produce offspring 1 and offspring 2 without repeated priority. They carry the genetic characteristics of the parent 1 and parent 2, and thus the crossover operation can guarantee the succession of the individuals.

3.5 Design of mutation operation

The mutation operation is to simulate the biological genetic and natural evolutionary process, where creatures' cell division and replication link may make replication errors because of some accidental factors, resulting in certain kind of mutation and thus new chromosomes with new characteristics of new biological traits. This paper mainly adopts the mutation based on a central location, which can be divided into four steps: ① calculate the number of mutated genes, $U = total\ number\ of\ genes \times pm$, U is the number of mutated genes, pm is the coefficient of mutation of the total group; ② generate a random number of U (betwteen1 and the total number of genes) as a variation of the gene; ③ locate the relevant chromosome; ④ adopt the method of mutation based on the central location, randomly exchange value with other alleles within the chromosome, and finally generate a new chromosome.

4. Application examples

This section first gives a resource-constrained instance, then apply the genetic algorithm with the model establish in this paper, and finally achieve the optimal project scheduling.

4.1 Background

The knitting machine is the main equipment of the producing sweaters. In the wool market, whether domestic or oversea market, the demand is very strong. After the financial crisis is over, Europe and the United States consumption market starts to recover, which leads to effective demand. With the ever-increasing demand, more businesses are put into the knitting machine production. But the knitting machine market in China has been long occupied by foreign products. The main reason behind the lack of such products in China is that the product development process is complex, demanding large capital investment and long development cycle, while the key to determine whether a company can survive lied in its production.

The population size of the project is 200, the variation rate is 0.2, crossover rate is 0.2, γ is 0.8, the hereditary algebra is 50, the number of resource one, two and three is respectively 6,4 and 2. Project network diagram is shown in Figure 5.

According to the mathematical model established in the third part, initialize the adjacency matrix in the project activity diagram, determine the fitness function, run the C + + programming according to the parameters set, and after 50 generations of iterative optimization, the optimization result is 600.

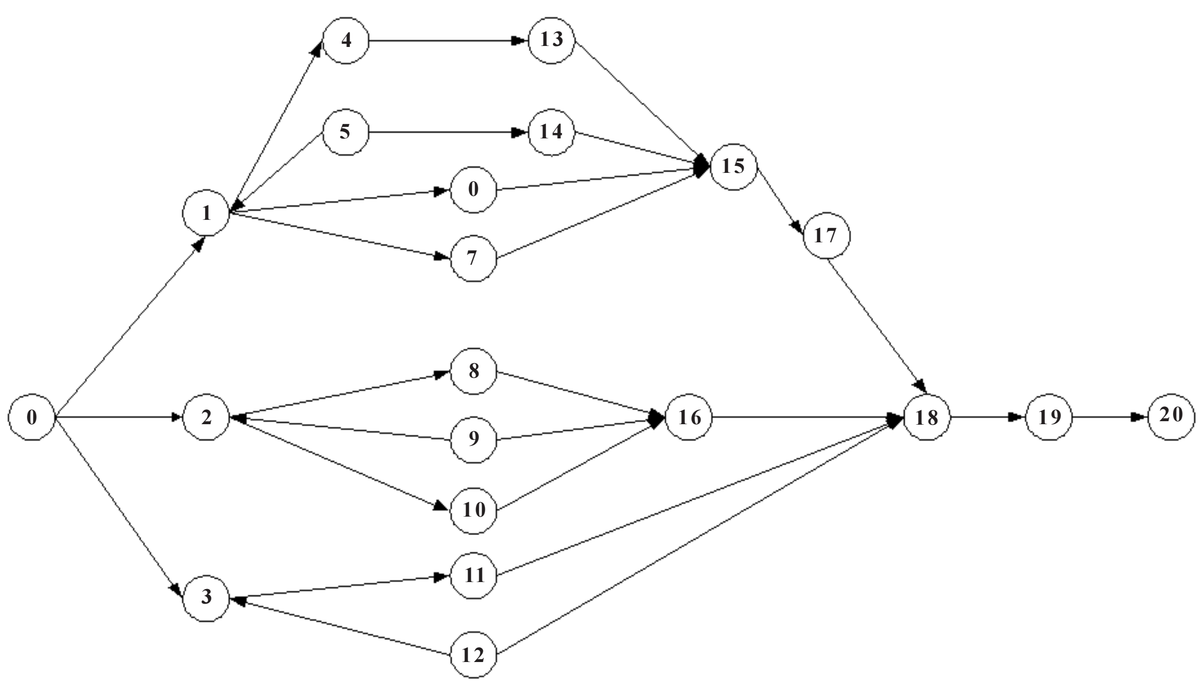

Figure 5. Project network diagram

The project activity list is shown in table 2.

D	Title of activity	Duration	Resource 1	Resource 2	Resource 3
0	Start	Virtual activity	3	0	0
1	Driver Board Section	100	3	0	0
2	Motherboard Section	100	3	0	0
3	PC Section	120	3	0	0
4	Needle Reading Section	80	2	2	1
5	Progressive Motor Section	50	2	3	1
6	Push Pin Triangle Design	40	2	3	1
7	Yarn Mouth Design	20	2	3	1
8	Servo System	40	3	3	1
9	Raab System	30	3	3	1
10	Rack Sensor	20	1	2	2
11	Pattern Preparation System	150	3	0	0
12	GUI Interface	50	3	0	0
13	Needle Selection Section	30	2	2	1
14	Stitch Section	30	2	2	1
15	Hill-shaped Plate Design	50	1	3	1
16	Rack Design	40	1	3	1
17	Head Design	20	1	3	1
18	Machine Design	60	3	4	2
19	The Whole Test	20	3	2	1
20	End	Virtual Activity	3	2	1

Table 2. Project activity information

Project optimization curve is shown in Figure 6. The figure indicates the average time of the project is slightly undulating, but with the increase of the genetic algebra, on the whole it shows an ever-decreasing tendency, which proves that the genetic algorithm has realized the optimization of project scheduling. There may be several reasons behind the fluctuations may be due to the following points:

(1) With the iteration process and the operation in the form of roulette, better fitness may be chosen into the next generation of groups, but due to the randomness of the whole operation, the possibility of poor individuals pouring into the genes of next generation can’t be ruled out, and thus relatively good individuals has not been selected.

(2) Since the possibility of cross and variation for each chromosome is possible in the next generation of groups, some individuals with excellent fitness may confront genetic crossover and mutation, and thus have worse fitness, which may lead to the rise of the average value of the entire group.

The optimal curve of the optimized time shows that with the increase of genetic algebra the fluctuations are comparatively obvious in the genetic medium period, and as the iterations continue the value gradually is stabilized at 600. Through analysis of the phenomenon, possible causes are the following:

(1) In the process of producing the following groups, a crossover happens to the individual corresponding to the optimal solution, so the individual of the optimal solution has been replaced by a new individual. But the target value of the new individual is greater than that of the previous generation of the optimal solution, which leads to the rise of the curve.

(2) In the process of producing the next generation, variation happens to the individual corresponding to the optimal solution, resulting in a target value that is greater

than that of the optimal solution and eventually the rise of the curve.

4.1.1 Parametric analysis

It can be predicted that when parameters such as the population size, genetic algebra, mutation rate and crossover rate change, the convergence of the genetic algorithm will also be affected.

1) This paper first takes the population size parameter as an example to explore the impact of population size on the convergence of the optimization process. When the population size is 100, the project schedule needs optimization. Clearly, compared to the number of the population in the initialization part of 200, the optimal solution value has changed a lot. When the number of the population is 200, the shortest time of the project is 600. When the population size is 100, the minimum optimization time of the project is 620. Accordingly it can be inferred that a larger population size is beneficial to get the optimal solution for the project. The optimization curve is shown in Figure 7.

2) On the basis of the analysis above, this paper will also take impact of the mutation probability on the final results into consideration. When the mutation rate dropped from 0.2 to 0.02, the volatility of the optimization curve is clearly reduced, and the convergence effect of the curve is also better than that when the mutation rate is 0.2. Therefore, it is not difficult to conclude that the decrease of the group variation rate is conducive to the stability of the entire group, and the values of the optimal scheduling time also indicate obvious decrease. However, it shall not be ignored that a much low mutation probability means that population is disabled of variation, so that when the genes of the initial population are not very good, the difficulty of the optimal solution of the population will be greatly increased. The optimization curve when the mutation probability is 0.02 is shown in Figure 8.

4.1.2 Performance comparison

Referring to the experimental results of genetic algorithm for project optimization scheduling in Reference [10], both the ever-decreasing average and minimum time of the final optimization and the convergence of the optimal value are better than the experimental effects of this paper's design. From the project optimization curve designed in this paper, it can be seen that there are repeated fluctuations for both average and optimal time, and the convergence effect is also not as good as that of the genetic algorithm in the reference. The optimal curve given in the reference is shown in figure 9.

It can be seen from the figure above that the stability of the optimize curve given in the reference is better, the average value and the optimal value shows stability and ever-decrease tendency with the increase of genetic algebra, and in later stage of the heredity the average value curve and the optimal value curve almost overlap, indicating that there is a high proportion of excellent individuals in the group and the convergence effect is also obvious.

Compared with other algorithm methods given in other references, it is not difficult to find that according to the algorithm method given in Reference [10], the individual of the optimal solution is reserved into the next generation of groups in the process of producing next generation of groups, without exchange and variation of chromosomes, which is exactly the reservation of genes. As a result, the optimal solution of the next generation is certainly not

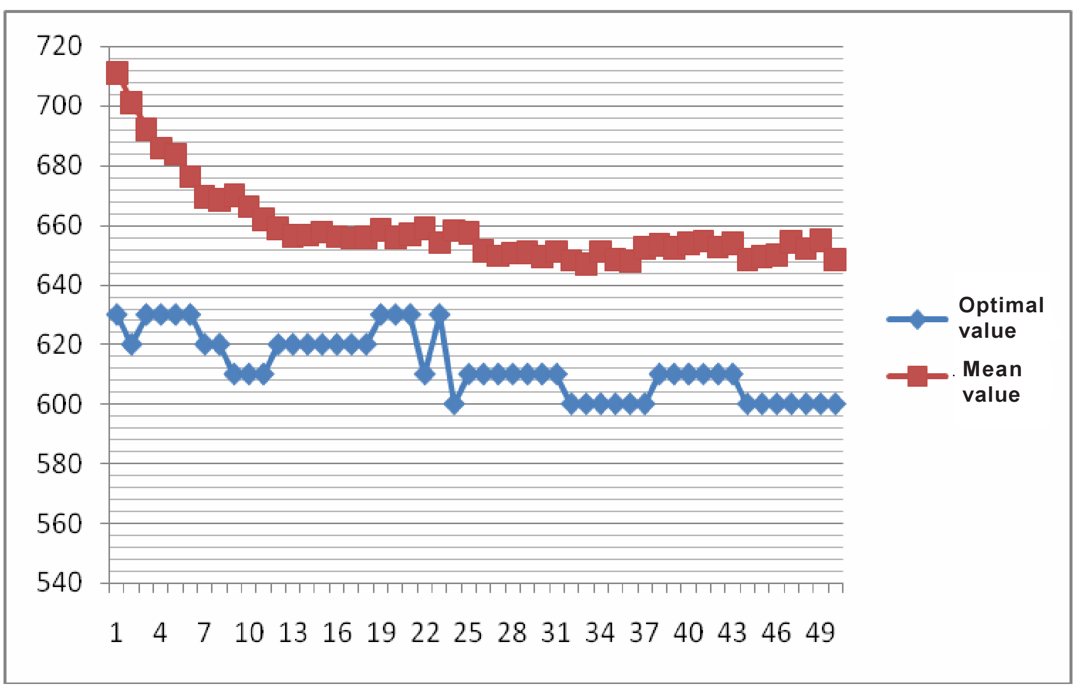

Figure 6. Project Duration Curve

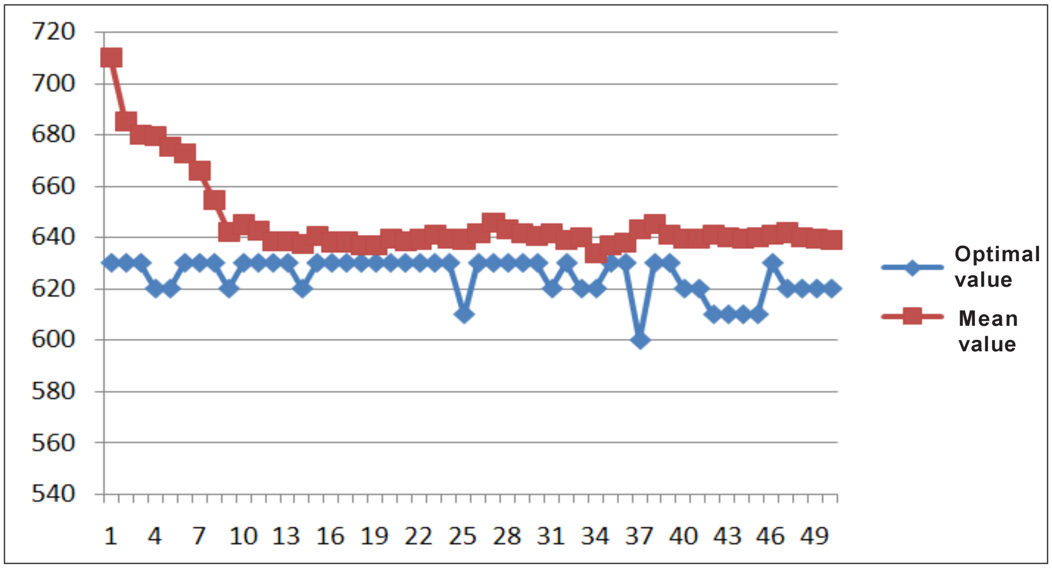

Figure 7. Optimization curve when the population size is 100

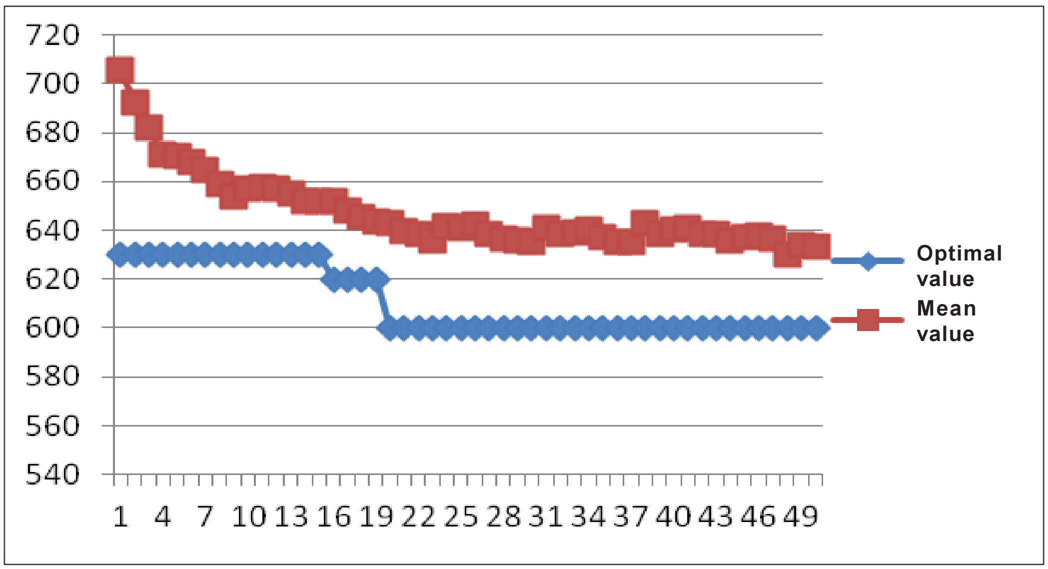

Figure 8.Optimization curve when the mutation probability is 0.02

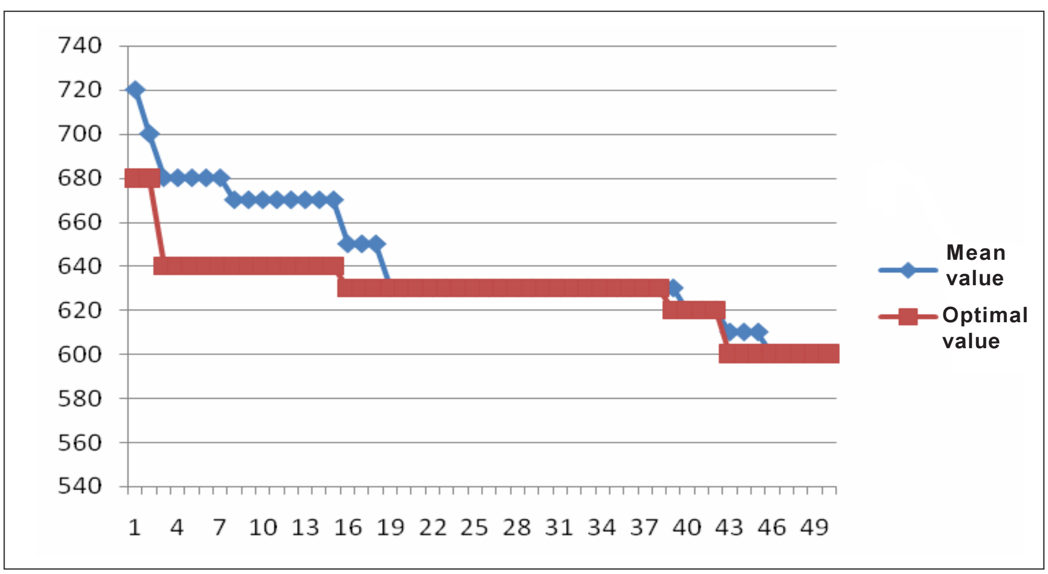

Figure 9. Comparison of the optimization Curves

larger than that of the previous generation, and thus the convergence of the entire group is enhanced, and the optimization results are improved.

6. Conclusions

In the market economy, it has become the concern of each enterprise in sense of product development to adequately and reasonably take advantage of the existing enterprise resources, speed up project development, push its R & D products to market in the shortest possible time and thereby gain market share. This paper is based on the genetic algorithm, and analyzes project scheduling optimization in case of resource constraints. The particular activities in the project, in line with the timing and with the number of resources available, can be executed in different time. It has a very important theoretical and practical significance to flexibly select the best execution time under the premise of satisfying the predecessor constraints relationship between the resources and activities, and finally achieving the shortest entire duration of the project.

7. Acknowledgement

This research is supported by the Natural Science Foundation of Zhejiang, China(Grant No. Y1111159).

References

[1] Masao, M., Ching, C. T. (1997). A genetic algorithm for multi-mode resource constrained project scheduling problem. *European Journal of Operational Research*, 100 (1) 134-141.

[2] Bouleimen, K., Lecocq, H. (2003). A new efficient simulated annealing algorithm for the resource constrained project scheduling problem and its multiple mode versions. *European Journal of Operational Research*, 149 (2) 268-288.

[3] Bowman, T., Edward, H. (1959). The Schedule sequencing problem. *Operations Research*, 18 (7) 621-624.

[4] Kelley, J. E. (1963). The Critical Path Method: Resource Planing and Scheduling. Englewood Cliffs: Pren2tice2Hall.

[5] Mohring, R., Schulz, A., Stork, F., Uetz, M. (2003). Solving Project Scheduling Problems by Minimum Cut Computations. *Management Science*, 49 (3) 330-350.

[6] Ning, Mao, Qing, Chen, Chen, Xin. (2001). An Exact Algorithm in MPCPSP. *Control Theory and Applications,* 18 (1) 119 - 126.

[7] Shou Yongyi. (2004). Resource Constrained Multiple Project Scheduling Iterative Algorithm. *Journal of Zhejiang University* (Engineering Science), 38 (8) 1095-1099.

[8] Xu, Cheng, Wu, Chengqing. (2006). A Complex Procedure of Hybrid Intelligent Algorithm for Project Scheduling Problem. *Computer Integrated Manufacturing Systems*, 12 (4) 585-589.

[9] Shixin, Liu, Jianhai, Song, Jiafu, Tang. (2006). Resource-Constrained Project Scheduling In The Buffer Set Method. *Systems Engineering*, 21(4) 381 - 386.

[10] Alcaraz, J., Maroto, C. (2001). A Robust Genetic Algorithm for Resource Allocation in Project Scheduling. *Annals of Operations Research*, 102 (124) 83-109.

Author Biography

Chunlai Chai is currently teaching at College of Computer Science and Information Engineering, Zhejiang GongShang University. He received Master Degree Computer Science Hohai Univerity,Nanjing,China,2003. He has published over 10 papers in international journals and conferences in the areas of wireless communications, E-Commerce and Data Mining.