World Transactions on Engineering and Technology Education
Vol.14, No.1, 2016

Optimal allocation of higher education resources based on the Pareto genetic algorithm

Wei He†, Pan Liang‡ & Guiying Zou†

ABA Teachers University, Chengdu, People's Republic of China†
Chengdu Aeronautic Vocational and Technical College, Chengdu, People's Republic of China‡

ABSTRACT: Higher education resources are among the scarcest of social resources, and the contradiction between demand for higher education and the shortage of higher education resources is high at present in China. In order to calculate the optimal solution of the allocation of higher education resources, the optimisation of manpower resources of higher education has been modelled using the Pareto genetic algorithm. It reduced the complexity of the model without designing personal preferences. According to the nature of problems, a hybrid multi-objective genetic algorithm with a high solving efficiency has been designed. It provides a technical method for the optimisation of the allocation of higher education resources.

INTRODUCTION

Higher education resources are among the scarcest of social resources, and the contradiction between demand for higher education and shortage of higher education resources is high in China. To solve the contradiction, other than by injecting more investment into higher education, it is important to ensure the maximum benefit through the rational allocation of current resources.

The authors of this article endeavoured to model the optimisation of manpower resources of higher education with the Pareto genetic algorithm to work out the Pareto optimal solution. They reduced the complexity of the model without designing personal preferences. According to the nature of problems, the authors designed a hybrid multi-objective genetic algorithm with a high solving efficiency. They adopted the related Pareto concept and effectively generated the Pareto solution. This provides a technical method for the optimisation of the allocation of higher education resources.

THE MULTI-OBJECTIVE PROGRAMMING PROBLEM AND PARETO SOLUTION SET

General Model of Multi-Objective Programming for Resources Allocation of Higher Education

In essence, the analysis of resources allocation for higher education can be viewed as a multi-objective programming problem.

Provided that $f(x) = (f_1(x), f_2(x),\ldots,f_n(x))$, it refers to the objective function vector of various higher education resources, $g(x) = g_1(x), g_2(x),\ldots,g_k(x))$ refers to the constraint function vector of all kinds of resources; $x = (x_1,x_2,\ldots,x_m)$ refers to all kinds of resources vectors.

One can convert this problem into the following programming model:

$$\max\{Z_1 = f_1(x), Z_2 = f_2(x),\ldots, Z_n = f_n(x)\}$$

$$s.t. g_i(x) \le 0, i = 1,2,\ldots,k$$
$$x = (x_1, x_2,\ldots, x_m), x_i \ge 0, i = 1,2,\ldots,m$$

The feasible region is expressed as:

$$Z = \{z \in R^n | z_1 = f_1(x), z_2 = f_2(x),\ldots, z_n = f_n(x), x \in S^m\}$$

where S^m refers to the feasible region of decision space, R^n refers to the feasible region of objective space.

Scholars at home and abroad proposed a strategy for the solution of multi-objective programming. In order to be easily understood, here are a few related concepts.

The Pareto Solution Set

In view of the maximisation problem of an objective function, one can define the dominant solution, the non-dominated solution, the Pareto optimum solution, the Pareto frontier and other concepts.

Definition 1 (dominant solution), if, and only if, there are decision variables $x_v \in S^m$ making: $\forall_1 \in \{1,2,...,n\}, v_i > u_i$,

Calling that $v = \{v_1,...,v_n\}$ dominates $|u = \{u_1,...,u_n\}$, this moment, x_v is called as the dominant solution of x_u, where $v_i = f_i(x_v), i = 1,...,n$ and $u_i = f_i(x_u), i = 1,...,n$.

Definition 2 (non-dominated solution), if, and only if, there are no decision variables $x_v \in S^m$ making:

$\forall_1 \in \{1,2,...,n\}, v_i$, calling X_u as the non-dominated solution.

Definition 3 (Pareto optimum solution and Pareto frontier), the set of non-dominated solution is called the Pareto optimum solution, denoted by PS,

$$P_F = \{ f(x) = (f_1(x), f_2(x),..., f_n(x)) | x \in P_s \}|$$

is called the Pareto frontier.

THE RELATIONSHIP OF SOLUTION SET OF MULTI-OBJECTIVE PROGRAMMING AND SOLUTION STRATEGY

The Relationship of Solution Set of Multi-objective Programming

Pareto first studied multi-objective optimisation problem in the field of economics and his theory is called the Pareto optimum theory. The so-called ideal solution produced by solving the single objective optimisation problems is often outside the feasible region. People are looking for the best solution in the single objective optimisation, while looking for a set consisting of the Pareto optimum solution [1] and non-dominated solutions [2] in multi-objective programming. The relationship diagram of multi-objective solution set is shown in Figure 1.

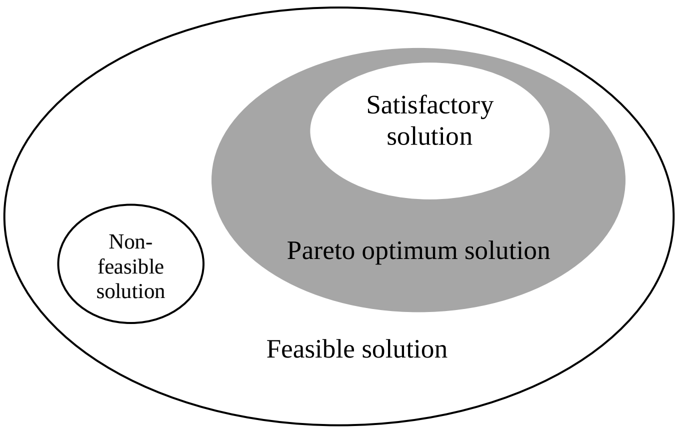

Figure 1: The relationship diagram of multi-objective solution set.

It can be seen from Figure 1 that each solution in the Pareto optimum solution set is a non-dominated solution of the multi-objective programming problem. The genetic algorithm can search multiple non-dominated solution through the population of the whole solution set in parallel way, and policymakers can choose the decision scheme in more than one of the solutions, which means a very attractive method for solving the multi-objective programming problem. In fact, what single objective programming solved is the optimum solution, and the same for multi-objective programming is the satisfactory solution.

Conventional multi-objective optimisation methods solve multi-objective optimisation problems, which usually refer to synthesising a multi-objective to a single objective for optimisation based on a utility function. However, in most cases, the defect in this method is: 1) the utility function before optimisation is difficult to determine and often only produces one solution; 2) a failure to ensure Pareto optimality unless having obtained the optimum value of objective function in advance; 3) the weight or preferences is not easy to get; 4) the dimensions between multiple objectives do not match and are not easy to handle; and 5) even through the variable weight method was used, it still fails to find the solution at the non-convex boundary through integrating the multiple objectives [3].

The main difference between a multi-objective optimisation problem and a single objective optimisation problem is that the solution of multi-objective optimisation is not the only one, but is an optimal solution set based on the elements, which are called Pareto solutions or non-dominated solutions.

The multi-objective evolutionary algorithm, based on the evolutionary mechanism, is a global optimisation search algorithm, which is able to solve complicated multi-objective optimisation problems. Due to adopting the concept of Pareto optimisation, a multi-objective genetic algorithm can approximate the Pareto frontier through the evolution of generations of population to overcome the limitations of traditional methods in solving the multi-objective problems. The multi-objective genetic algorithm has become a very successful application (e.g. production scheduling, sorting, etc) in many aspects of combinatorial optimisation. Using the Pareto optimum solution can overcome the problem of optimum solutions obtained by conventional methods, which could not be compared with each other.

Goldberg proposed the fitness function Pareto method based on the rank, the basic idea of which is directly mapping multiple objective values to the fitness function and, then, on the basis of Pareto optimum set, obtaining the fitness function form based on the rank; firstly, by forming a vector with the multiple objective function values to represent an individual, provided that while the individual X_i is in the generation t and dominated by $P_i(t)$ individuals, then, their rank in the generation is: rank $(X_i, t) = 1 + P_i(t)$.

As shown in Figure 2, the rank of the whole non-dominated individuals are 1, individual 3 is inferior to individual 2, because the individual 2 is in the compromise area. For the calculation of fitness, it notes that not all rank exist in some generations. There is no rank 4 in Figure 2, so the method of computing according to the rank must be adjusted. Firstly, sorting the individuals in the population and, then, calculating the individual fitness according to the linear interpolation method, the number 1 will be the optimum individual.

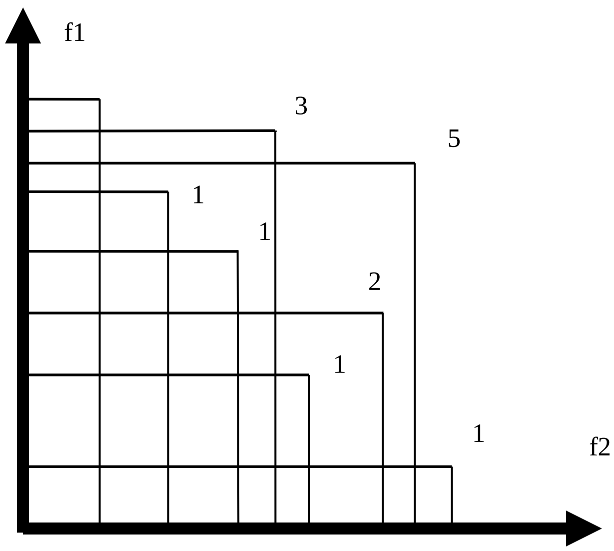

Figure 2: The Pareto rank of individuals in the population.

Using the average fitness of individuals with the same rank as the fitness of the combination population to ensure that the fitness value is a constant will guarantee that the distribution of fitness is uniform. However, a large amount of calculation is required to solve the rank, especially for larger populations, and the fitness function is a single function, which is easy to cause the search carried out in a particular direction and the genetic drift.

Looking at the literature, Fonseca and Fleming proposed the multi-objective genetic algorithm with fitness assignment based on sorting [4]; Horn and Nafpliotis proposed the niche Pareto genetic algorithm [5]; Srinivas and Deb proposed the genetic algorithm with sorting the non-dominated solutions [6]; Knowles and Corne proposed the archiving evolution strategy of Pareto solutions [7]; and Zitzler et al proposed the Pareto concentration evolutionary algorithm, and proposed the improvement of the Pareto concentration evolutionary algorithm [8]. These algorithms overcame the problem of larger population size and solved the problem that it is easy to cause the search carried out in a particular direction and the genetic drift as the fitness function is a single function.

THE ESTABLISHMENT AND ANALYSIS OF OPTIMAL ALLOCATION MODEL FOR HIGHER EDUCATION RESOURCES

The Combinatorial Optimisation of Production Elements for Higher Education

The optimisation allocation system of higher education resources is a complex, multidimensional and non-linear system. The optimum combination of production factors is the factors' combination form, which refers to producing the required product at the lowest possible cost. To cultivate more students with the least amount of manpower, material resources and financial resources for higher education:

Provided that Z is the education output and $x_i (i = 1,2,3,\ldots,n)$ is the different combinations of education elements, then: $Z = f(x_1, x_2,\ldots,x_n)$.

If the combination of education elements is different, the education output is also different, and there must be an optimal combination among the whole combination, which refers to a minimum input, but a largest output. How to calculate the optimal combination, as the Western education economy theory said, depends mainly on the production efficiency of education; namely, the ratio between the education input and output.

$$P = \sum_{j=1}^{m} Z_j / \sum_{i=1}^{n} x_i$$

where P is the ratio between input and output; x_i is the different elements combination; Z_j is the output.

The larger P is, it means the more reasonable combination of education production factors, while the smaller P is, it means a more unreasonable combination; hence, the combination with the largest input-output ratio is the optimal combination. However, an increase in output may result in an increase in education cost. Therefore, the cost issues need to be considered. In this article, the authors synthetically considered the cost factor, established a multi-objective optimisation model and found satisfied satisfactory solution with the Pareto genetic algorithm.

The optimisation of higher education resources has no uniform standard, but it has a unified goal. As universities shoulder three big functions, which refer to teaching, scientific research and social service, three types of indices: mean number of students per faculty, per capita output of scientific research and per capita income of social service, can be set, respectively, to reflect the efficiency of human resources in teaching, scientific research and social service. For material resources, it will be the total fixed assets at the end of the year, which also actually reflect the quality of students; for financial resources, it will be education expenditure; and for the output of products, it will be the number of students in school, funds raised for science and technology activities or topic quantity, etc.

Finally, having synthetically considered quantity and quality factors, the following six operability indices were determined to be used: A_1 - the proportion of students and teachers (%), A_2 - the proportion of teachers accounted for all the staff (%), A_3 - per capita number of teachers' scientific research projects, A_4 – the average fixed assets of students at the end of the year (thousand yuan), A_5 – the average education expenditure of student (thousand yuan), and A_6 - the number of students in school (ten thousand people). There have been some scholars at home and abroad using the above index values to calculate the composite score of university resources utilisation efficiency. The calculation formula is:

$$S = C_1 * \underline{A}_1 + C_2 * \underline{A}_2 + C_3 * \underline{A}_3 + C_4 * \underline{A}_4 + C_5 * \underline{A}_5 + C_6 * \underline{A}_6 .$$

The determination and calculation method of indices here refers to Wang [9]. Where $\underline{A}_i$ is the score of A_i, C_i is the weight (i = 1,…,6). Due to the subjectivity of the weight, the dimensions between indices are different. So, the authors performed regularisation processing for some indices, established a model with multi-objective programming theory and found a satisfied solution with the Pareto genetic algorithm. However, it should be noted that multi-objective decision-making problems usually have the following characteristics:

- The contradiction of objectives: objectives have contradictory phenomena in multi-objective decision-making. For example, if one wants to increase the number of university students, one should increase funds investment, teachers and equipment, and may even reduce the freshman admission criteria, etc, but these objectives are contradictory.
- Objectives are non-commensurable: in general, objectives have completely different nature and measurement unit; for example, the six indices have no nature of commensurability.
- The decision results are affected by the preferences of decision-makers: decision-makers have different attitudes to risk or different preferences for some objectives, which will influence the result of the decision, as mentioned above, and the weights of six indices will be influenced by personal preferences.

$$\max R_1 = \sum_{i=1}^{n_j} x_{ji} / \sum_{i=1}^{n_i} y_{ji} \quad (1)$$

$$\max R_2 = \sum_{i=1}^{n_j} y_{ji} / \sum_{i=1}^{n_i} z_{ji} \quad (2)$$

$$\max R_3 = \sum_{i=1}^{n_j} D_{ji} / \sum_{i=1}^{n_i} y_{ji} \quad (3)$$

$$\max R_4 = \sum_{i=1}^{n_j} E_{ji} / \sum_{i=1}^{n_i} y_{ji} \quad (4)$$

$$\max R_5 = \sum_{i=1}^{n_j} F_{ji} / \sum_{i=1}^{n_i} x_{ji} \quad (5)$$

$$\max R_6 = \sum_{i=1}^{n_j} x_{ji} \quad (6)$$

$$s.t.\ \alpha_{11} \le \sum_{i=1}^{n_j} x_{ij} \le \alpha_{12}, \forall j \quad (7)$$

$$\alpha_{21} \le \sum_{i=1}^{n_j} y_{ij} \le \alpha_{22}, \forall j \quad (8)$$

$$\alpha_{31} \le \sum_{i=1}^{n_j} y_{ij} \le \alpha_{32}, \forall j \quad (9)$$

$$\alpha_{41} \le \sum_{i=1}^{n_j} y_{ij} \le \alpha_{42}, \forall j \quad (10)$$

$$\alpha_{51} \le \sum_{i=1}^{n_j} y_{ij} \le \alpha_{52}, \forall j \quad (11)$$

$$X_{uv}, Y_{uv}, D_{uv}, E_{uv}, F_{uv}, H_{uv} \ge 0 \forall, uv \quad (12)$$

This model was established aiming at the indices of regional universities: R_1 - the proportion of students and teachers, R_2 - the proportion of teachers accounted for all the staff, R_3 - per capita number of teachers' scientific research projects (R&D), R_4 - average fixed assets of students at the end of the year, R_5 - average education expenditure of students, R_6 - number of students in school.

In the model, n_j is the number of universities in region j, provided that n_j refers to the number of universities countrywide, then, this model can be used for the resources optimisation of universities countrywide.

$\sum_{i=1}^{n_j} U_{ij}$ refers to the annual total of the corresponding index in region j. These indices are respectively: total number of students at school in the year (U refers to X), total number of full-time teachers in the year (U refers to Y), total number of academic and general staff in the year (U refers to Z), total number of scientific research projects in the year (U refers to D), total value of fixed assets in the year (U refers to E), total value of education expenditure in the year (U refers to F).

The objective functions are: 1) maximisation of the proportion of students and teachers; 2) maximisation of the proportion of teachers accounted for all the staff; 3) maximisation of the per capita number of teachers' scientific research projects; 4) maximisation of the average fixed assets of students at the end of the year; 5) minimisation of the total value of education expenditure; and 6) maximisation of the number of students in school.

Constraint (7) - (11) refers to the boundaries that ensure corresponding indices of a school in line with the actual situation (a_{i1} and a_{j2},i = 1,…,5 and are the lower bound and upper bound of resources). Equation (12) refers to a larger variable value than the zero restriction.

SOLUTION OF THE MODEL

First designing coding methods, fitness assignments, genetic operators, constraint processing, etc, which can be able to effectively solve the problem must be carried out. The main method of hybrid multi-objective evolutionary algorithm is as follows. Using the selection and elitism to the initial population randomly generated to identify the good chromosomes and obtaining new chromosomes with crossover and mutation operation, each of which corresponds to a Pareto solution, approaching the Pareto frontier through continuous evolution [10].

Coding

For example, as shown in Figure 3, there is the form of a chromosome, which adopts the decimal coding with a total of six.

0.21	0.22	0.16	0.13	0.18	0.1

Figure 3: The composition of a chromosome.

Among them, each number is a decimal, which refers to the effect size of corresponding index in the optimal allocation of resources.

Fitness Function

The essence of optimal allocation for higher education resources is using resources to create maximum efficiency; therefore, the objective function can be used as fitness function. The quick non-dominated sorting method proposed by Deb is used to solve the non-dominated solution set [11].

The main idea is to compare every solution of the population with other solutions to get a Pareto solution set, according to the concept of Pareto solutions. Such a solution set is called the first Pareto frontier, denoted as F_1. By lie over the first Pareto frontier, perform the comparison according to the concept of Pareto solutions continually to get the second Pareto frontier, and so on, one can obtain the third and fourth Pareto frontier, etc; finally, one can get the Pareto frontier in the sorting $\{F_1, F_2, \ldots\}$.

Selection Operation

Adopting the competition way to select chromosomes, the method is as follows: one should randomly select two individuals from the population, denoted as P_1 and P_2. Then, compare P_1 and P_2, and choose the Pareto frontier individual in dominant; if P_1 and P_2, and are in the same Pareto frontier, the one less crowed should be chosen.

Crossover and Mutation Operations

Due to the possible violations to model constraint for the new solution, crossover and mutation operations must be feasible; namely, ensuring that each operation makes each chromosome satisfy the constraints. Using a single-point crossover to randomly select two chromosomes from current population and one point on the basis of hybrid rate P_c to directly cross. A single-point mutation was used to select a mutation point, on the basis of mutation rate P_m, to mutate for the chromosomes in current population. In order to ensure that mutated chromosomes satisfy constraints, the variation method of increasing or decreasing offset was used.

Pareto Genetic Algorithm Process of Model

The algorithm process is as follows:

Step 1: Perform the regularisation processing to the objective value of individual.
Step 2: Set the evolution algebra M, population size N, crossover rate P_c and mutation rate P_m. Generate an initial population P_0 with $t = 1$, $P_t = P_0$.
Step 3: Calculate the fitness value on the basis of individual Pareto rank and share the operation.
Step 4: Perform a selection, crossover and mutation operation to get a new population.
Step 5: Perform the quick non-dominated sorting method to the new population with a P_{t+1} generated.
Step 6: Judge whether the end condition $t = t + 1$ is met, if not, turn to Step 4; otherwise, end with a result output.

ANALYSIS OF SIMULATION RESULTS

Using VC++ programming for the simulation calculation and the processing method is divided into two steps:

First of all, regularisation processing was performed to multiple objectives to combine them into one objective with a combination method and calculate the optimum combined coefficient with general genetic algorithm, which refers to recombining R_i to make:

$$S = \sum_{i=1}^{6} C_i V(R_i),$$

This results in the following models:

$$\begin{cases} \max S = \sum_{i=1}^{6} C_i V(R_i) \\ s.t. \sum_{i=1}^{6} C_i = 1; C_i \geq 0, i = 1,2,3,4,5,6 \\ \gamma_i \leq R_i \leq \beta_i, i = 1,2,3,4,5,6 \end{cases}$$

Where $\sum_{i=1}^{6} C_i = 1$ with the same significance of C_i ditto, $V(R_i), i = 1,2,3,4,5,6.$ is the objective function after normalisation, γ_i, R_i, β_i respectively refer to lower bound and upper bound of each R_i.

Calculate the optimum C_i, i = 1,2,3,4,5,6, with general genetic algorithm, which results in:

C_1 = 0.143, C_2 = 0.137, C_3 = 0.177, C_4 = 0.144, C_5 = 0.261, C_6 = 0.138, with a fitness of 0.90.

Thus, the biggest factor that influences resources allocation efficiency is not the proportion of students and teachers, but the average education expenditure per student, followed by the per capita number of teachers' projects, while the others are basically the same.

According to the original multi-objective programming model, let all the parameter values randomly change in the possible range, obtain the Pareto solution of index value with Pareto genetic algorithm, when the genetic algebra is 10000, one gets the result:

$$R_1 = 17.212, R_2 = 61.587, R_3 = 0.736, R_4 = 46.70, R_5 = 12.16, R_6 = 57.021.$$

The optimum ratio of students to teachers is 17:1; the optimum proportion of full-time teachers is 61.5%; the optimum per capita number of teachers' scientific research projects (R＆D) is 0.7; the optimum average fixed assets of students at the end of the year is 46,700 yuan; the optimum average education expenditure on students is 12,200 yuan; and the optimum number of students in school (region) is 570,000. Along with the ongoing evolutionary process, the solution obtained continuously approaches the Pareto frontier. The data have no fundamental changes with genetic algebra increasing to 50,000. Putting these data into the model generated by combination with a fitness of 0.96 is obtained. It has no difference with the fitness above.

The authors carried out a comprehensive score to the actual data of various regions in recent years [12-13] with the above optimal solution. The score formula is:

$$S = C_1 * \underline{A}_1 + C_2 * \underline{A}_2 + C_3 * \underline{A}_3 + C_4 * \underline{A}_4 + C_5 * \underline{A}_5 + C_6 * \underline{A}_6,$$

where the significance of (C_i and $\underline{A}_i, i = 1,...,6$) is the same as above, the difference is that the C_i is no longer subjectively determined, the calculation method of $\underline{A}_i$ is statistically grouping the indices according to the range scope of corresponding index values, sorting them from low to high, respectively, giving the rating score from high to low with 100, 90, 80, … and, then, calculating the regions scores with comprehensively scoring formula to obtain the scores calculated as shown in Table 1.

Table 1: Scoring table of resources allocation for higher education in different regions.

Region	Score	Region	Score	Region	Score
Beijing	85.1	Anhui	92.2	Chongqing	83.2
Tianjin	86.4	Fujian	88.3	Sichuan	85.1
Hebei	89.3	Jiangxi	84.5	Guizhou	80.3
Shanxi	84.1	Shandong	86.2	Yunnan	80.4
Neimenggu	84.2	Henan	86.1	Xizang	78.1
Liaoning	88.6	Hubei	84.8	Shanxi	87.3
Jining	88.6	Hunan	87.8	Gansu	84.6
Heilongjiang	86.3	Guangdong	89.2	Qinghai	74.5
Shanghai	89.3	Guangxi	84.6	Ningxia	77.2
Jiangsu	90.0	Hainan	81.9	Xinjiang	77.5
Zhejiang	93.3				

The resources allocation efficiency of higher education in Shanghai, Jiangu, Zhejiang, Anhui, Fujian, Guangdong and Hebei is higher, and it is generally lower in the western regions. However, the relatively poor middle developed regions are Jilin, Hubei and Chongqing. It can be seen that the resources allocation is not singly dependent on the economy's development.

The authors found that the efficiency has improved with the increasing resources by changing the resources inputs. Due to the discreteness of the relationship curve between resources and efficiency, only the point forming the Pareto frontier has the nature.

CONCLUSIONS

This article presents a new attempt to study the optimal allocation of resources for higher education with a genetic algorithm. The most important contribution of this article is to realise the Pareto optimisation calculation and analysis of resources allocation for higher education with the Pareto genetic algorithm.

REFERENCES

1. Chankong, V.H., *Multi-Objective Decision Making Theory and Methodology*. New York: North-Holland, 5-6 (1983).
2. Hans, A.E., *Multi Criteria Optimization for Highly Accurate Systems*. New York: Plenum Press, 46-48 (1988).
3. Chankong, V.H., Op. cit. 56-58.
4. Fonseca, C.M. and Fleming, P.J., Genetic algorithm. Proc. Fifth Inter. Conf., San Mateo, America, 21-26 (1993).
5. Horn, J. and Nafpliotis, N., *A Niched Pareto Genetic Algorithm for Multiobjective Optimization. Proc. First IEEE Conf. on Evolutionary Computation, IEEE World Cong. on Computational Intelligence.* Piscataway, NJ: IEEE Service Center, 20-21 (1994).
6. Srinivas, N. and Deb, K., Multi-objective optimization using non-dominated sorting in genetic algorithms. *Evolutionary Computation*, 2, **3**, 221-248 (1995).
7. Knowles, J. and Corne, D., *The Pareto Archived Evolution Strategy: a New Baseline Algorithm for Pareto Multiobjective Optimisation. Proc. 1999 Cong. on Evolutionary Computation*. Piscataway,NJ: IEEE Press, 55-58 (1999).
8. Zitzler, E., Laumann, S.M. and Thiele, L., Improving the Strength Pareto Evolutionary Algorithm, Technical Report, 103, 1-21 (2001).
9. Wang, P., *Economics of Higher Education*. Beijing: Higher Education Press, 78-80 (2004).
10. Srinivas, N. and Deb, K., Op. cit.
11. Deb. K, Amrit, P. and Sameer, A., A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans on Evolutionary Computation*, 62, 182-197 (2002).
12. Ministry of Education, Statistical Yearbook of Education Funds in China. Beijing: China Statistics Press, 1-615 (2015)
13. Ministry of Education, Chinese Education Statistical Yearbook. Beijing: People's Education Press, 1-665 (2015)