Journal of Kerbala University , Vol. 7 No.1 Scientific. 2009

Applied in Assignment Problem Solving

Dhmyaa A. AL-Nasrawy
Ashwan A. AL-Ali
Asia.M.AL-Zubidy

Abstract

This paper attempts to implement survival of fittest by several selection strategies to solve the assignment problems through genetic algorithm . The assignment problem is basically the "N *men*- N *tasks*" problem where a single job can be assigned to only one person in such a way that the overall cost of assignment is minimized . Diversity has been measured of each generation and analyzed for a number of popular selection algorithms(Rolette Well Selection , Random Selection,and Tournament Selection with 4,8 size).Partial Matched Crossover(PMX) with mutation have been used with experiment crossover probability $P_c = 1$ and mutation probability $P_m = 0.5$.

المستخلص :

في هذا البحث نحاول تنفيذ مبدأ البقاء للأصلح بواسطة عدد من طرائق الانتقاء لحل مسالة الإحلال باستخدام الخوارزمية الجينية ، وبصورة أكثر تفصيل تعد مسالة الاحلال من المسائل المعروفة (N من المهام وN من الأشخاص) حيث يتم إحلال مهمة واحدة فقط لشخص واحد أو تكليف شخص واحد فقط لانجاز مهمة واحدة بحيث أن كلفة الإحلال تكون بأقل مايمكن . أيضا يتم قياس مدى التنوع الحاصل لكل جيل وتحليل النتائج بالنسبة لكل طرائق الانتقاء المستخدمة حيث تم استخدام طريقة PMX للتزاوج وباحتمالية = 1، وطريقة الإبدال للطفرة واحتمالية = 0.5 .

1. Introduction

Genetic algorithms are search methods that employ processes found in natural biological evolution. These algorithms search or operate on a given population of potential solutions to find those that approach some specification or criteria. To do this, the algorithm applies the principle of survival of the fittest to find better and better approximations. Selection operator is one of the important aspects in the genetic algorithm processes. A selection scheme in genetic algorithm is simply a process that favor the selection of better individuals in the population for the mating pool. The primary concern of all selection schemes is what's known as the loss of diversity. In a generational genetic algorithm (GA), not every individual makes it into the intermediate population. As a result, information encoded in the current population is not transferred into the next generation in its entirety. There are several ways for selection. Some of them are Tournament selection, Random selection, and Proportional selection. The Assignment Problem is one of the most famous problems in linear programming and in combinatorial optimization. There are a number of *men* and a number of *tasks*. Any man can be assigned to perform any task , incurring some *cost* that may vary depending on the man-task assignment. It is required to perform all tasks by assigning exactly one man to each task in such a way that the *total cost* of the assignment is minimized. If the numbers of men and tasks are equal and the total cost of the assignment for all tasks is equal to the sum of the costs for each man (or the sum of the costs for each task, which is the same thing in this case), then the problem is called the *Linear assignment problem* .Informally speaking, we are given an $n \times n$ cost matrix $C = (cij)$, and we want to select n elements of C, so that there is exactly one element in each row and one in each column, and the sum of the corresponding costs is a minimum [1] .

Mathematical model by introducing a binary matrix $X = (x_{ij})$ such that:

$$x_{ij} = \begin{cases} 1 & \text{if row } i \text{ is assigned to column } j, \\ 0 & \text{otherwise,} \end{cases}$$

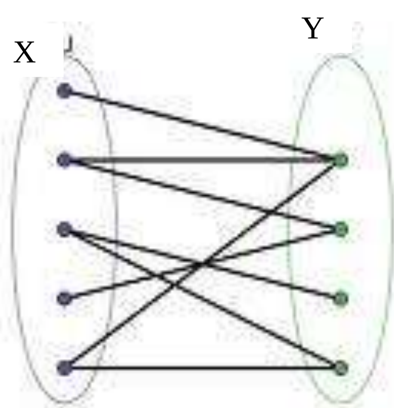
can be modeled as

$$\begin{aligned} \min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{n} x_{ij} = 1 \qquad (i = 1, 2, \dots, n), \\ & \sum_{i=1}^{n} x_{ij} = 1 \qquad (j = 1, 2, \dots, n), \\ & x_{ij} \in \{0, 1\} \qquad (i, j = 1, 2, \dots, n). \end{aligned}$$

2. Graphs for Assignment Problem

A graph is ordered pair G=(V,E) consisting of a finite set and a subset E of elements of the form (x,y) where x and y are in V. The set V are called the vertices of the graph and the set E are called the edges . A graph G is sad to be Bipartite (or Bicolored) graph if the vertices can be partitioned into two mutually disjoint sets X and Y so that there is no edge between any vertices in X or any vertices in Y. A bipartite graph will be denoted by G=({X,Y}, E) and represented by matrix. The X vertices are for example used for row indices and the Y vertices are used as column indices. The assignment problem corresponds to Bipartite Graph .The vertices V may be partitioned into two sets: the assigned tasks and the assignees . The edges consist of (unordered) pairs connected the assigned tasks and the assignees. Since we allow the theoretic possibility of assigning any task to any assigns , the graph for the assignment problem consists of vertices V=(X ,Y), where X={ $x_1, \dots, x_n$ } and Y={ $y_1, \dots, y_n$ }and edges consisting of all combinations E= { $(x_i , y_i) \mid 1 \le i \le n , 1 \le j \le n$ }. We can denote the incidence matrix of this graph as an $n \times n$ matrix consisting entirely of 1's and we can use the matrix as a substitute for

the graph ,figure(1) show example of bipartite graph[2] .



Figure(1) Example of Bipartite Graph

The Hungarian method used to solve the assignment problem ,also can be formulated as an integer programming model and solved by techniques such as Branch and Bound techniques . This paper attempt to solve this problem and evaluate the resulting solutions contained in generations with several selection strategies and measure the diversity in each generation[1].

3. Solving Assignment Problem in Genetic Algorithm

Genetic algorithms are search methods that employ processes found in natural biological evolution. These algorithms search or operate on a given population of potential solutions to find those that approach some specification or criteria. To do this, the algorithm applies the principle of survival of the fittest to find better and better approximations. At each generation, a new set of approximations is created by the process of selecting individual potential solutions (individuals) according to their level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics. This process leads to the evolution of populations of individuals that are better suited to their environment than the individuals that they were created from, just as in natural adaptation .

The main steps of a GA can be summarized as follows:
Initialize population P(0);

1. Evaluate each chromosome in the population;
2. WHILE "stop criteria" not satisfied DO
    1. Selection;
    2. Crossover;
    3. Mutation;
    4. produce the new population;
    5. Evaluate chromosomes of the population.

A basic genetic algorithm comprises three genetic operators(Selection , Crossover , Mutation). Starting from an initial population of strings (representing possible solutions), noted that the population was initialed randomly depending on permutation , each permutation represent one solution , example : when L=5 , 3,5,1,4,2 is solution , mean the first person take task three, and second person take task five ,and so on … . For each solution the cost was calculated upon cost matrix which also generated randomly . The GA uses these operators to calculate successive generations. First, pairs of individuals of the current population are selected to mate with each other to form the offspring, which then form the next generation[3] . We can easily conjecture that in case of N *men*, N *task*s, the total number of solutions possible is N! and our task is to select the best string (the one with minimum total cost). The initial population can be generated by permutations of a string with length L .
Some details about these operators in the following[4]:

3.1 Selection

This operator selects the chromosome in the population for reproduction. The more fit the chromosome, the higher its probability of being selected for reproduction. Thus, selection is based on the survival-of-the-fittest strategy, but the key idea is to select the better individuals of the population, as in tournament selection, where the participants compete with each other to remain in the population. The most commonly used strategy to select pairs of individuals is the method of roulette-wheel selection, in which every string is assigned a slot in a simulated wheel sized in proportion to the string's relative fitness. This ensures that highly fit strings have a greater probability to be selected to form the next generation through crossover and mutation. The best two individuals being selected replaced in worst two in the next generation , where the reminder population will be copied from old population[5] . There are various methods that can be used to select solutions out of the population. In the following discussion, it will be assumed that we are interested in a minimization problem. Therefore, instead of choosing solutions with a larger fitness, we will be interested in solutions with a lower COST. In addition, it is assumed that the solutions have been ordered in terms of increasing COST, so that the first solution has the lowest COST (highest fitness) and is the best solution in the population. After selection of the pairs of parent strings, the crossover operator is applied to each of these pairs [5] .This paper used genetic algorithm include the following types of selections

1- Roulette Well Selection : A selection operator in which the chance of a chromosome getting selected is proportional to its fitness (or rank). This is where the concept of survival of the fittest comes into play.

2-Tournament Selection: To run a tournament selection, a tournament size (*NTRN*) is chosen. The procedure is simply to randomly choose *NTRN* solutions from the population and select the one with the lowest COST (highest fitness). If this is done to create a Current Population for the next generation, the selected solution should be removed from the merged populations so that it cannot be selected again. Conversely, if this is used to create a Mating Population, the selected solution should remain since you want very good solutions to be chosen more than once.

3- Random Selection: A selection operator which randomly selects a chromosome from the population.

The properties of the survived child will then be used for future generation of the population where the desirable properties will be maintained .However, if a child is form with undesirable properties, it is most likely that the child will be eliminated and not be used for future generation. Thus the less desirable properties are removed from future generations. The ultimate aim of GA will be to have the average fitness of the population to increase with each generation under the constrained environment.

In cost terms, it would mean to have the cost of the parameters that is being considered to decrease with each generation. In order to find the most desirable solution to the constrained environment, the principle of survival of the fittest is used. Only the fittest out of the population will remain and be presented as the solution to the constraint problem [3] .

3.2 Crossover

Crossover is a genetic operator that combines (mates) two chromosomes (parents) to produce a new chromosome (offspring). The idea behind crossover is that the new chromosome may be better than both of the parents if it takes the best characteristics from each of the parents [3,6]. Crossover occurs during evolution according to a user-definable crossover probability. Here simple crossover will not work, because it result repetition that mean more person take the same task in our problem; instead we choose the method of Partial Matched Crossover(PMX) with crossover probability $P_c = 1$, to avoid the repetition **.**

3.3 Mutation

Mutation is a genetic operator that alters one or more gene values in a chromosome from its initial state. This can result in entirely new gene values being added to the gene pool. With these new gene values, the genetic algorithm may be able to arrive at better solution than was previously possible. Mutation is an important part of the genetic search as help helps to prevent the population from stagnating at any local optima[6] . Mutation occurs during evolution according to a user-definable mutation probability. Mutation applied by choosing two random locations in a string and swapping the corresponding values at that position with mutation probability $P_m = 0.5$.

4. Computing the Diversity

The simplest definition of diversity comes from the answer to the question "how different is everybody from everybody else?" If every chromosome is identical, there is no difference between any two chromosomes and hence there is no diversity in the population. If each chromosome is completely different from one another, then those differences add, and the population should be maximally diverse. So the diversity of a population can be seen as the difference between all possible pairs of chromosomes within that population. While the above definition makes intuitive sense, there is one aspect not covered: what do we mean by different. If a pair of chromosomes is only different by one locus, it only seems reasonable that this pair should not add as much to the diversity of the population as a pair of chromosomes with every locus different. Consequently the difference between chromosomes can be seen as the Hamming distance or chromosome distance, where the Hamming distance is the sum of all loci where the two chromosomes have differing genes. Hence the population diversity becomes the sum of the Hamming distances between all possible pairs of chromosomes. In cluster theory this is called the *statistic scatter* . Now, since the Hamming distance is symmetric, and is equal to 0 if the strings are the same, only the lower triangle in a chromosome-pairs table need be considered when computing the diversity. Consequently the all-possible-pairs diversity can be formalized as:

$$Div(P) = \sum_{i=1}^{n} \sum_{j=1}^{i-1} hd(ci - cj) \qquad \dots (1)$$

where *P* is the population and chromosome ci and cj $\in$ P and n is the population size and hd(ci-cj) is the humming distance between two chromosome [7] .

4. Applications and Results

In this paper, applying the genetic algorithm with each selection methods (Rolette Well Selection, Random Selection , and Tournament Selection with size 4 and 8) on Assignment Problem . The array of cost will be generated randomly at first time .The results yields with two experiment , first with small population size =20 ,and the second with large population size =50 . In two cases ,used the following parameters :

- Chromosome Length =20
- Number of Generation =10
- Type of Crossover = PMX with P_c= 1
- Type of Mutation = Exchange with P_m=0.5

First Experiment (Small Population size) : measured (diversity & average of fitness) in each generation with three selection methods ,with mutation and without mutation , see tables 1,2,3,and 4 , figures 2,3, 4 and 5 .

Table(1)Diversity without mutation (Small Population Size)

Gen.	Diversity			
	RWS	Rand	Tour-4	Tour-8
1	3585	3556	3561	3581
2	3573	3499	3524	3511
3	3556	3509	3408	3381
4	3539	3500	3395	3300
5	3521	3460	3340	3144
6	3508	3334	3280	2901
7	3426	3332	3159	2590
8	3415	3178	3049	2194
9	3388	3036	2946	1850
10	3399	2758	2865	1425

Table(2)Diversity with mutation (Small Population Size)

Gen.	Diversity			
	RWS	Rand	Tour-4	Tour-8
1	362	355	3561	3594
2	359	349	3496	3490
3	359	343	3435	3436
4	358	327	3410	3276
5	355	304	3322	3231
6	357	296	3153	3123
7	356	266	3066	2926
8	354	240	2896	2708
9	354	210	2714	2425
10	355	172	2496	2109

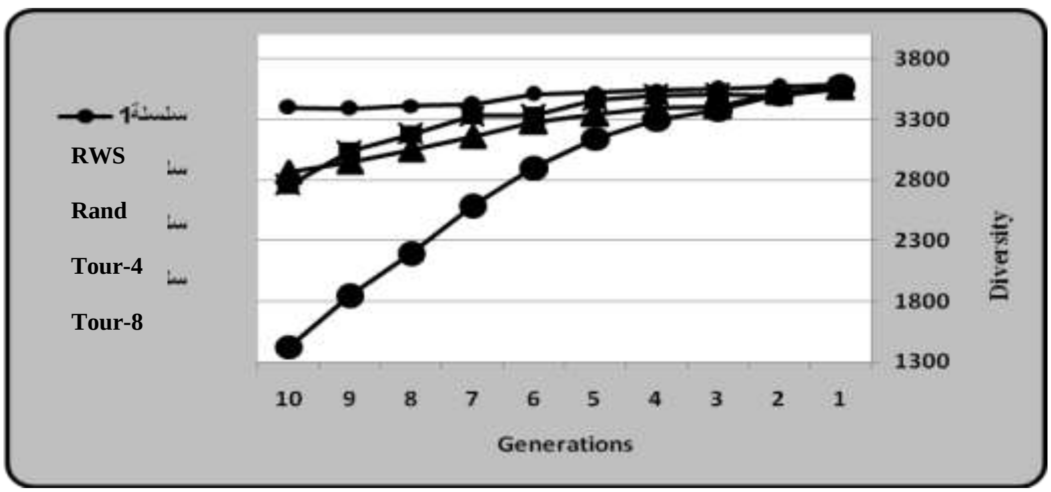


Fig (2)Diversity without mutation (Small Population Size)

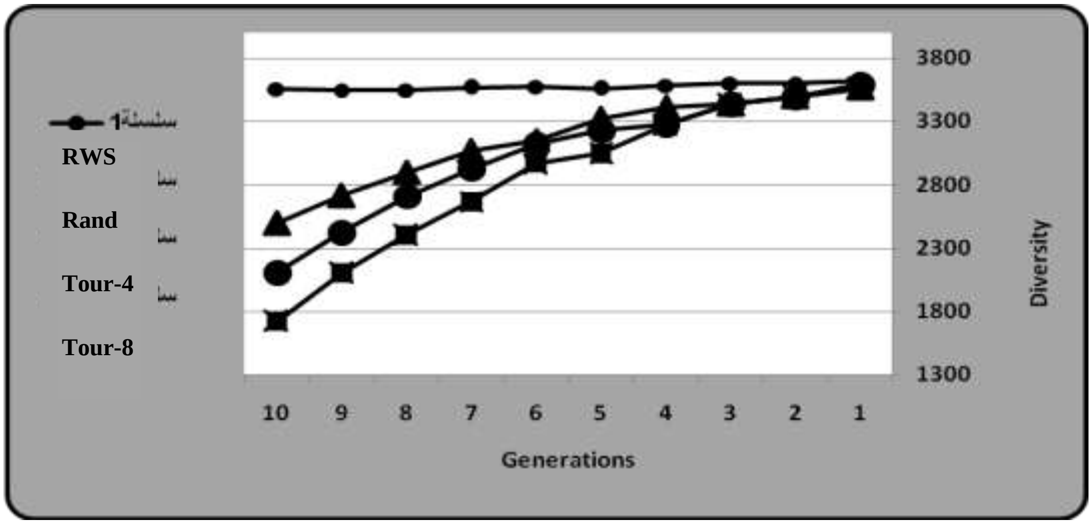


Fig (3)Diversity with mutation (Small Population Size)

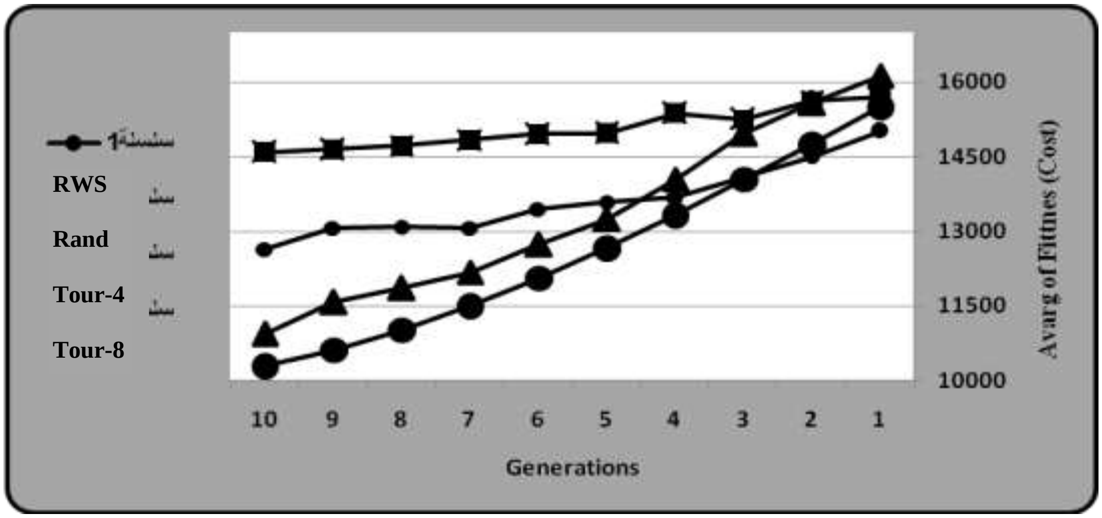
As showing from fig(2),fig(3), the diversity increased when using mutation process, and the RWS result high measure of diversity ,but the other methods of selection suffered from loss of diversity with and without mutation .

Table(3) Average of Fitness (Cost) without mutation(Small Population Size)

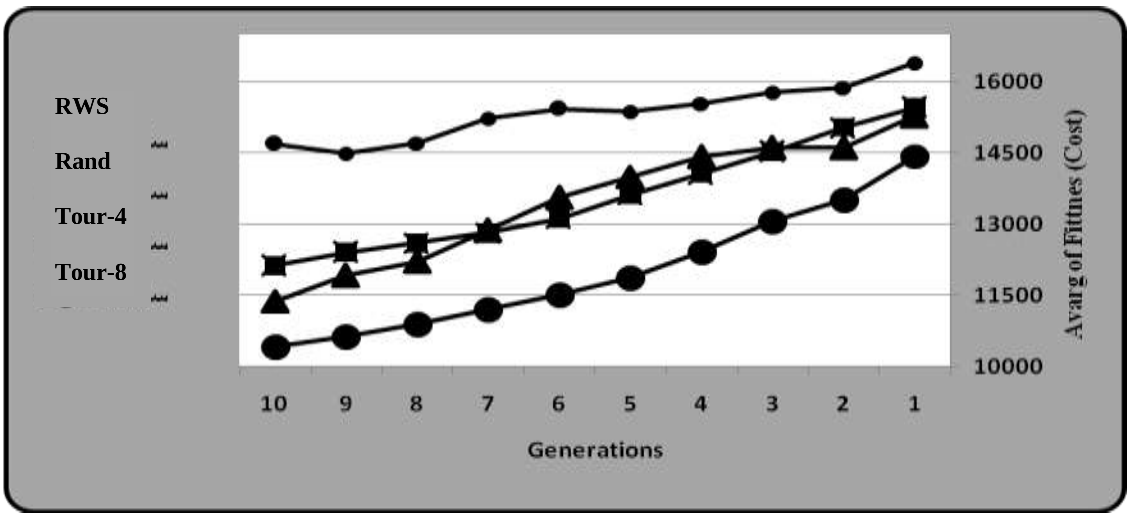
Gen.	Cost			
	RWS	Rand	Tour-4	Tour-8
1	15013	15689	16136	15497
2	14494	15623	15578	14754
3	14081	15252	14952	14050
4	13679	15371	14051	13328
5	13582	14970	13241	12659
6	13444	14955	12730	12059
7	13062	14838	12175	11495
8	13091	14720	11860	11016
9	13061	14646	11563	10609
10	12635	14594	10923	10289

Table(4) Average of Fitness (Cost) with mutation(Small Population Size)

Gen.	Cost			
	RWS	Rand	Tour-4	Tour-8
1	16397	1546	15283	14436
2	15877	1504	14616	13511
3	15783	1453	14606	13071
4	15539	1406	14422	12429
5	15367	1362	14009	11869
6	15433	1311	13559	11515
7	15232	1282	12881	11205
8	14703	1261	12220	10906
9	14496	1240	11929	10641
10	14695	1213	11390	10418



Fig(4) Average of Fitness (Cost) without mutation(Small Population Size)



Fig(5) Average of Fitness (Cost) with mutation(Small Population Size)

The Tour-4 ,Tour-8 methods of selection result minimal cost that mean high average of fitness ,see fig(4),fig(5).

Second Experiment (Large Population size) : measured (diversity & average of fitness) in each generation with three selection methods ,with mutation and without mutation , see tables 5,6,7, and 8 ,figures 6,7, 8 and 9 .

Table(5)Diversity without mutation (Large Population Size)

Gen.	Diversity			
	RWS	Rand	Tour-4	Tour-8
1	23290	23340	23319	23185
2	23261	23303	23252	23053
3	23230	23265	23221	23042
4	23212	23264	23169	22996
5	23169	23218	23054	22915
6	23132	23086	22950	22756
7	23118	23144	22715	22554
8	23091	23150	22641	22241
9	23059	23090	22327	21827
10	23016	23011	21922	21351

Table(6)Diversity with mutation (Large Population Size)

Gen.	Diversity			
	RWS	Rand	Tour-4	Tour-8
1	23312	23279	23308	23342
2	23281	23255	23283	23276
3	23272	23213	23229	23249
4	23251	23139	23150	23211
5	23248	23080	23096	23105
6	23211	23084	22980	22996
7	23163	23031	22916	22806
8	23122	22958	22858	22752
9	23101	22801	22675	22616
10	23097	22834	22630	22284

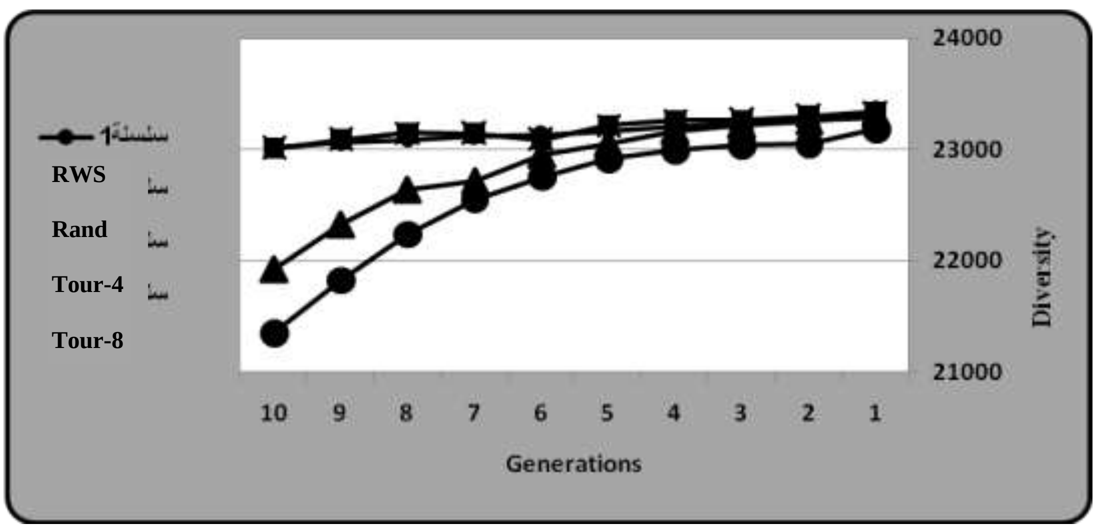


Fig (6)Diversity without mutation(Large Population Size)

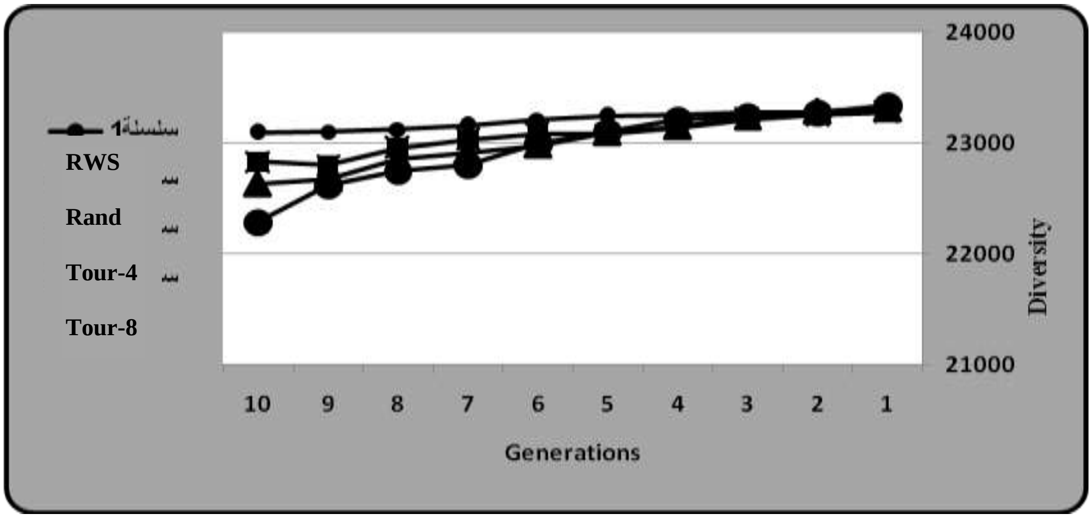


Fig (7)Diversity with mutation(Large Population Size)

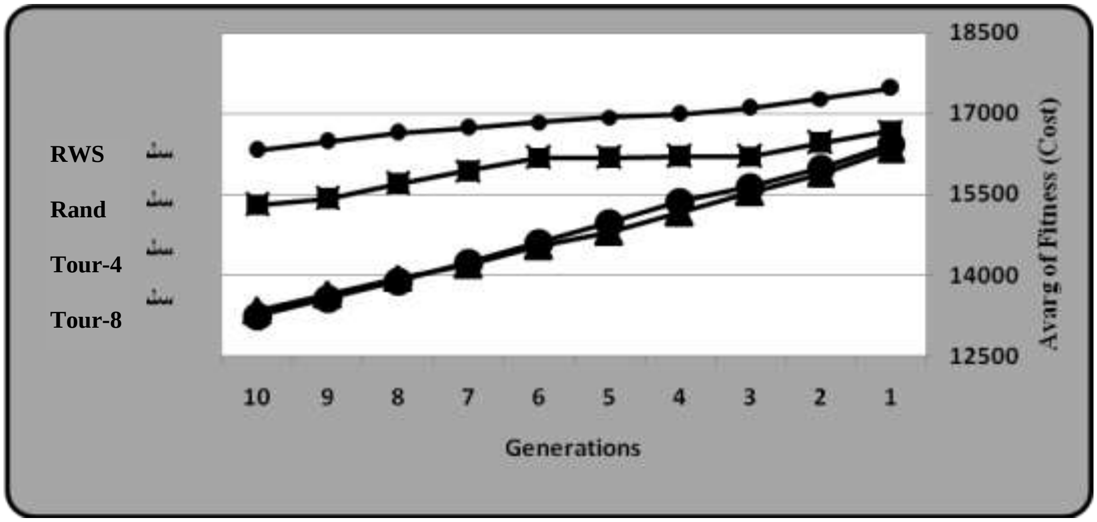
Good result obtained in large population size with mutation when applying the selection methods specially RWS method , show fig(6),fig(7) .

Table(7) Average of Fitness (Cost) without mutation(Large Population Size)

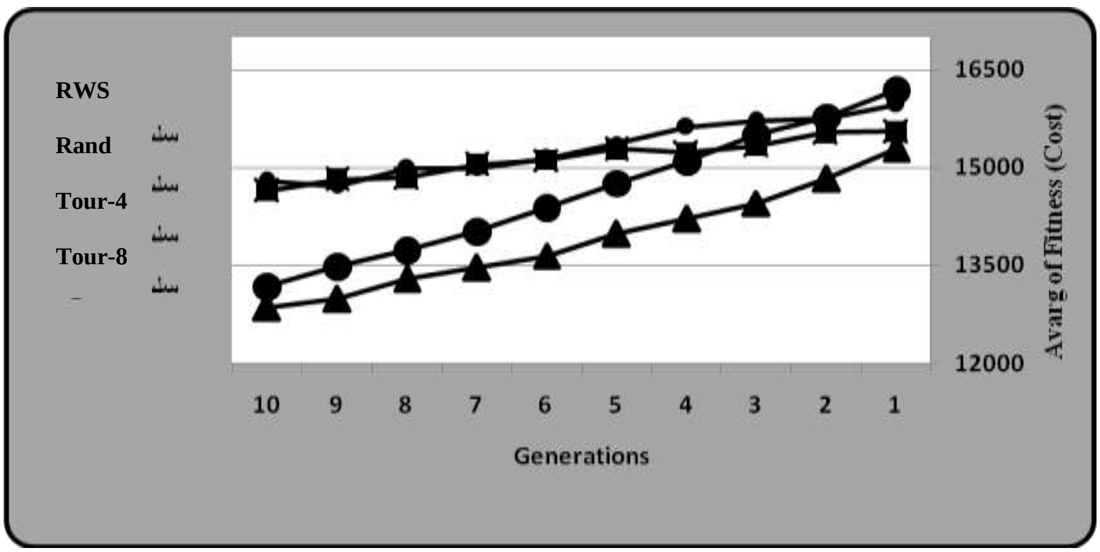
Gen.	Cost			
	RWS	Rand	Tour-4	Tour-8
1	17469	16681	16328	16436
2	17273	16458	15885	15986
3	17105	16200	15542	15639
4	16993	16205	15155	15360
5	16929	16182	14801	14972
6	16842	16179	14536	14600
7	16732	15939	14199	14245
8	16644	15697	13952	13898
9	16486	15425	13631	13566
10	16324	15301	13337	13245

Table(8) Average of Fitness (Cost) with mutation(Large Population Size)

Gen.	Cost			
	RWS	Rand	Tour-4	Tour-8
1	15970	15553	15288	16192
2	15746	15533	14832	15777
3	15719	15339	14456	15498
4	15619	15227	14208	15104
5	15364	15283	13986	14759
6	15145	15117	13647	14383
7	14987	15051	13466	14024
8	14989	14845	13290	13738
9	14722	14818	12991	13488
10	14793	14647	12852	13174



Fig(8) Average of Fitness (Cost) without mutation(Large Population Size)



Fig(9) Average of Fitness (Cost) with mutation(Large Population Size)

Minimal cost obtained to solving assignment problem when using Tour method of selection with two size (4,8) , with and without mutation process , see fig(8) , fig(9) . The differences between small and large population size explained in the result of diversity in above figures .

5. Conclusion

The selection procedure chosen, and the values of any parameters used, can significantly affect the performance of the algorithm and the results obtained. If you want an algorithm that converges to a result rapidly, you should choose a method that selects only the better solutions in the population. In this paper ,the Tour selection method give us minimal cost in small population size, but increase the loss of diversity .The four types of selection methods apply the principles of " Survival for Fittest", but not all of these method gives optimal solution. Also noted that the RWS and large size of population give high measure of diversity specially with using the mutation .

6. References

[1] Billy E. Gillett, *Introduction to Operations Research A Computer-Oriented Algorithmic Approach.* Tata McGraw-Hill Publishing Company Limited, New Delhi (Copyright © 1979 by McGraw-Hill, Inc., New York), TMH EDITION 1979, ch. 3, ch.4.

[2] S.Baase , Computer Algorithm , Introduction to Design and Analysis , Addison Wesely , Publishing Company ,1988

[3] Omar Al Jadaan, Lakishmi Rajamani,C. R. Rao ,Improved Selection Operator for GA , 2008.

[4] Anshuman Sahu, Rudrajit Tapadar , solving the Assignment problem using Genetic Algorithm and Simulated Annealing , IAENG International Journal of Applied Mathematics, 36:1, IJAM_36_1_7 ,2007 .

[5] A.Townsend , Genetic Algorithms – a Tutorial ,2003 .

[6] Goldberg, D.E., Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, Mass.: Addison-Wesley, 1989.

[7] Mark Wineberg , Franz Oppacher , The Underlying Similarity of Diversity Measures Used in Evolutionary Computation .