Probability model to Solve the School Bus Routing Problem with Stops Selection

Ricardo Pérez-Rodríguez, Arturo Hernández-Aguirre
Center for Mathematics Research CIMAT, AC
ricardo.perez@cimat.mx

Abstract. This paper describes the use of a new algorithm that solves the school bus routing problem with stops selection. The aim is to reduce the travel time of a set of buses that transport students to a school. The fundamental contribution of the authors is the use of a probability model to describe the feasible solution space distribution and thus to get the best solution. An estimation of distribution algorithm is used to address the combinatorial complexity of the problem statement. Different and better results are obtained with the proposed algorithm against a genetic algorithm. The contribution of this paper is to propose an alternative to solve permutation-based representation problems with logistics application.

Keywords: probability model; school bus routing; vehicle routing; genetic algorithm; estimation of distribution algorithm.

1 Introduction

The transportation of students is an important factor in the efficient management of the services offered by the different schools especially when such transportation complexity increases as the number of students to attend. A large number of transport companies and schools in different countries are faced with the transport situation every day for students. Various laws that regulate the way in which students are transported and to where the bus stops can be placed is discussed, approved and modified continuously. The school bus routing problem is a logistics issue that requires more accountability and better service by drivers and the companies that offer the service. Bus stops are constantly changing. As students move or change of degree, it may be necessary to move, add or delete any stop on the route. Security considerations also come into play with the location of stops, for example, whether the bus turns to the right for a pickup or students crossing the street can be critical in making decision the correct location of the stop bus. An additional requirement may be that a stop should be located at a distance of no more than n meters in the direction of each student.

Unlike traditional vehicle routing formulations where there is a set of nodes (vertices) already established, this research focuses on the school bus routing problem where a set of potential bus stops is known. Then, determining the set of bus stops really to use is a part of the formulation in this research. The objective of this problem is (1) finding the set of bus stops to visit, (2) determine for each student what bus stop should be addressed and (3) set the path for the chosen bus stops, so that the total travel time is minimized.

Minocha and Tripathi [1] describe a real school bus routing problem. The authors develop a roadmap for the bus service of a school located in Rajasthan, India, so it is able to serve students in an efficient manner with maximum capacity utilization of the buses using a hybrid genetic algorithm. In this research, the authors transform the school bus routing problem with the format of a vehicle routing problem with time windows. Therefore, the vertices of the graph are the stops where students board the bus. Schittekat et al. [2] solve a school bus routing problem with stops selection using a meta-heuristic. The meta-heuristic proposed called adaptive randomized greedy search procedure, based on a method of neighborhood, use an exact algorithm to solve optimally the problem of allocation of students to the stops when the routing is defined. This meta-heuristic starts from a solution in which all stops are visited on separate routes. Therefore, the solution is represented by numbers representing the buses. The results of the meta-heuristic approach was compared with solutions found by a sequential method, with solutions obtained by applying a mixed integer programming model using commercial software, and solutions obtained by a column generation approach.

Niu [3] focuses on how to determine an operations schedule for an urban transportation company (a real case) for the peak hours during the morning. An operations schedule at irregular intervals intends to match the demands in terms of time and increase the use of vehicles. This program allows buses depart at times different from each other. In addition, a linear programming model is formulated to minimize the waiting time and costs incurred by the use of vehicles. Finally, a heuristic method based on a genetic algorithm is developed to generate feasible solutions. The representation of the solution is binary and indicates that a bus part or not in the corresponding time point.

Diaz-Parra et al. [4] proposed an application of a bio-inspired algorithm for a school bus routing problem with the features of a single school, urban services, mixed cargo, special students and homogeneous fleet. In this study, each chromosome is represented by a set of numbers. Each number represents a bus stop. The order of the numbers shows a feasible route. A set of test cases were developed by the authors and these cases were resolved by the proposed genetic algorithm.

Other contributions that are virtually in the same direction of current research are Euchi and Mraihi [5], Pacheco et al. [6] and a review of the problem mentioned can be found in Park and Kim [7]. Most of the methods developed for the school bus routing problem with stops selection are characterized by using a sequential approach [2], that is, a selection procedure for the bus stops is followed for a routing process or vice versa. In the first case (selection-routing), it has the drawback of generating more routes than the opposite case (routing-selection). Moreover, the trend of current research is to develop algorithms and iterative search procedures where the process of variation (represented by the cross and mutation operators on the algorithms based-population) has control of the solutions that are generated without having totally clear for the user as such solutions were obtained. The random nature of the process of variation makes changes to the existing solutions as a black box approach, whereas in this study the black box approach is avoided, that is, the search process is transparent.

The Estimation of Distribution Algorithms (EDA) is a relatively new paradigm in the field of evolutionary computing. Compared with other evolutionary algorithms, in the EDA the new population is reproduced without using traditional evolutionary operators. In the EDA, a probability model of the most promising area is built with statistical information based on the search experience, and then the probability model is used to generate new individuals. The EDA make use of a probability model to describe the solution space distribution. The updated process of the solution space distribution reflects the tendency of the evolution population (Wang et al. [8]). To describe the solution space distribution, the EDA tries to determine a relationship or interaction between the variables of the problem as its main objective. Traditional evolutionary operators are replaced by a probability model that is built with information about relationships and interactions between the variables of the problem. The main idea is to learn and benefit from interaction between variables by estimating the population distribution and sample new offspring from it. Although the interaction may or may not be present, usually, this is explicitly unknown even for the school bus routing problem. In this investigation, an EDA is proposed in order to improve the sequential selection-routing approach. Finally, although different algorithms have been proposed in order to solve the above problem successfully, in the best of our knowledge, probability models have not been used to address the problem described.

2 Experimental procedures

2.1 Allocation

As the goal in this research is to improve the sequential approach for building solutions for the school bus routing problem with stops selection known as selection-routing, then the first decision is to assign students to the bus stops by means of an iterative process that assigns each student to a bus stop in the first possible stop within the set of potential stops and the number of students assigned to the bus stop mentioned is updated to avoid exceeding the capacity of any bus. A stop is considered potential when a student walk to the stop and the distance-time restriction is not violated. Therefore, a stop may be potential for a student, but it is not potential for other students. The code iterative procedure described below in Figure 1.

```
Define distance-restriction
Define bus-capacity

for each student i do
for each stop j do
distance = (stop location j, student location i)
if distance ≤ distance-restriction
      feasible matrix (i,j) = 1
else
      feasible matrix (i,j) = 0
endif
endfor
endfor

for each stop j do
capacities(j) = 0

for each student i do
assigned = 1
for each stop j do
      if feasible matrix (i,j) == 1 AND capacities(j)<bus-capacity
          assigned = j
            break
      endif
endfor
capacities(assigned)=capacities(assigned)+1
endfor
```

Fig. 1. A method of allocating student-stop.

As an example, Schittekat et al. [2]'s first test is used, which has 5 stops with 25 potential stops to assign students, and where the distance-time restriction is only 5 min. Then given the above procedure a feasible allocation is below

1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4 5 5 5 5 5

Where each number represents the stop assigned to each student.

The second decision concerns the routing procedure. For this research a population of M vectors is initialized with size m, where m is the total number of selected stops. Each vector is initialized as a permutation of selected stops mentioned. A permutation is shown below according to the above example

4 2 5 1
Where the first bus stop to attend is the 4 after the 2 and so on.

Once the permutation is generated, a feasible solution should be created so that the capacity of the vehicles is not exceeded based on the Prins [9] procedure. This is for each vector of the population.

2.2 Selection

A subset N of M parents is selected according to their fitness value by the method of the bubble. Fitness in this case is the total travel time.

2.3 Probability model

The proposed probability model uses the simplest approach for estimating the joint probability distribution of selected individuals in each generation, $p_t(x)$. This joint probability distribution is factored as a product of the univariate marginal distributions $p_t(x)=p(x|N)=\prod^l p_t(x_l)$. Each univariate marginal distribution is estimated with marginal frequencies $p_t(x_l)=\sum^N \delta_j(X_i=x_i|N)/N$ where: $\delta_j(X_i=x_i|N)=1$, if the j-th case N, $X_i=x_i$

Figure 2 details an example where for the first position (first column) in the permutation are only six possible stops according to the individuals selected. Therefore, the possible values to choose for the first bus stop on the route (first column) are 2, 3, 4, 6, 7 and 9 and each has a 1 /6 chance of being chosen.

2.4 Sampling

To generate new offspring, each bus stop is selected with probability pj at each position of the route according to the selected individuals, see Figure 2. If a bus stop has already been chosen, it means that the bus stop cannot be selected again. Then, for each successor position of the route should not be considered the bus stops selected previously. This update mechanism considers previous assignments. Figure 3 depicts for the first position the bus stop 4 was selected. Then the bus stop 4 cannot be elected for the second position. The possible values to choose for the second position in the route (second column) are 5 and 7 and each has ½ chance of being elected.

Sequence position

X1	X2	X3	X4	X5	X6	X7	X8
2	6	8	3	4	5	9	7
3	4	5	9	7	2	6	8
9	7	2	6	8	3	4	5
6	8	3	4	5	9	7	2
4	5	9	7	2	6	8	3
7	2	6	8	3	4	5	9

Probability model 1st position X1

Probability model 2nd position X2
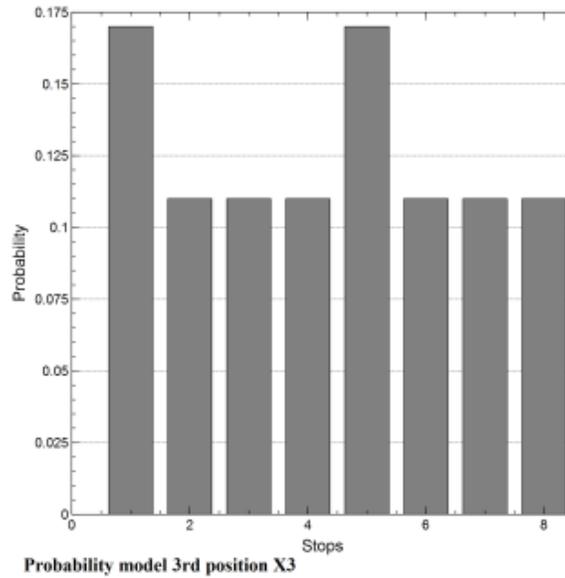

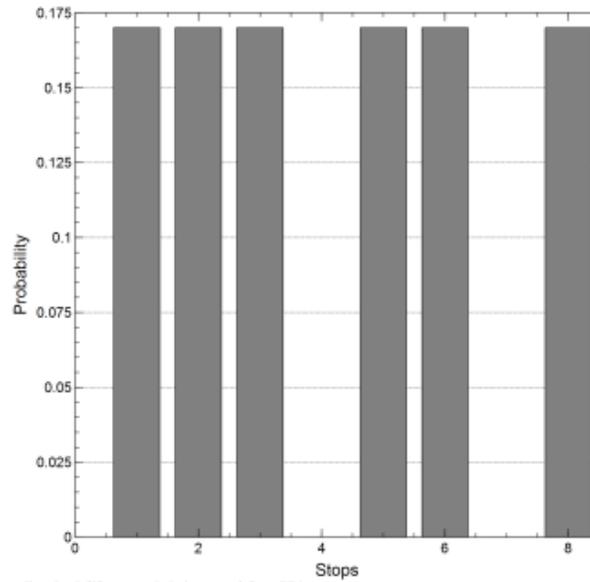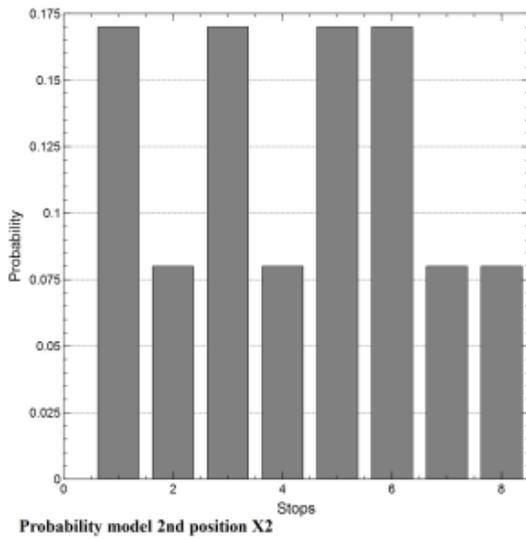Probability model 3rd position X3

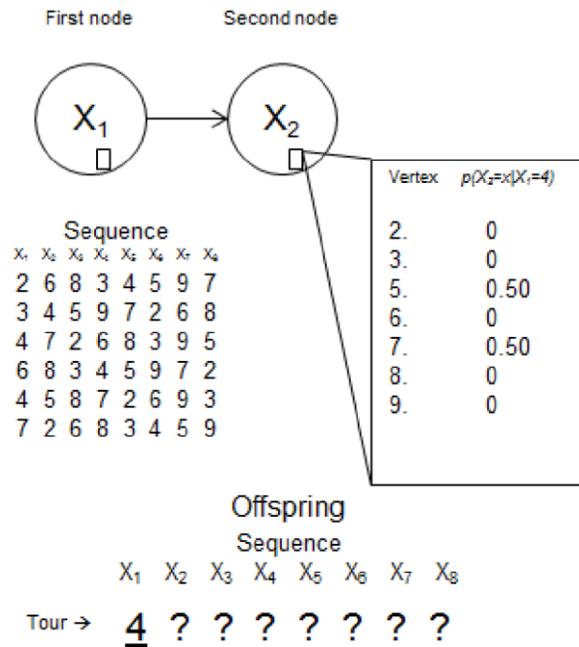**Fig. 2.** Probability distribution generated for each bus stop.

Fig. 3. Update on the probability distribution of the sample.

3 Results

Schietekat et al. [2]'s test sets serve as input parameters to show that the sequential selection-routing approach can be improved. A better position to each bus stop in the sequence can be estimated by the probability model. The tests mentioned ranging from 5 stops and 25 students to 800 students and 80 stops. In addition, four maximum walking distances are considered: 5, 10, 20 and 40. The maximum walking distance-time determines the average number of stops that a student is able to walk. From Table 1 to Table 5 include seven columns: the instance (id column), the number of stops (stop column), the number of students (stud column), bus capacity (cap column), the maximum distance-time walk (column wd), the shortest travel time obtained with a genetic algorithm used as comparative (GA column), and the proposed EDA (EDA column). Table 1 provides details on the results for 5 potential stops. From Table 1 can see how the probability model is able of detecting competitive sequences than the genetic algorithm for the problem stated.

Table 1. Comparative results for 5 potential stops

id	stop	stud	cap	wd	GA	EDA
1	5	25	25	5	140.99	140.99
2	5	25	50	5	161.31	161.31
3	5	25	25	10	181.27	181.27
4	5	25	50	10	194.01	194.01
5	5	25	25	20	133.30	133.30
6	5	25	50	20	106.00	106.00
7	5	25	25	40	36.77	36.77
8	5	25	50	40	47.15	47.15
9	5	50	25	5	285.80	285.80
10	5	50	50	5	197.61	197.61
11	5	50	25	10	203.99	203.99
12	5	50	50	10	214.64	214.64
13	5	50	25	20	129.89	129.89
14	5	50	50	20	94.39	94.39
15	5	50	25	40	19.95	19.95
16	5	50	50	40	30.86	30.86
17	5	100	25	5	361.17	361.17
18	5	100	50	5	302.20	302.20
19	5	100	25	10	295.74	295.74
20	5	100	50	10	237.61	237.61
21	5	100	25	20	155.90	155.90
22	5	100	50	20	158.53	158.53
23	5	100	25	40	**59.98**	75.77
24	5	100	50	40	**46.74**	55.28

Table 2 provides information on the results when there are 10 potential stops. From Table 2 can see how the probability model is able to detect competitive sequences and sometimes getting less time than the genetic algorithm.

Table 2. Comparative results for 10 potential stops

id	stop	stud	cap	wd	GA	EDA
25	10	50	25	5	253.18	**251.92**
26	10	50	50	5	332.09	**307.79**
27	10	50	25	10	275.34	**271.74**
28	10	50	50	10	325.27	**299.04**
29	10	50	25	20	109.82	109.82
30	10	50	50	20	195.99	**193.32**
31	10	50	25	40	33.25	33.25
32	10	50	50	40	67.21	67.21
33	10	100	25	5	401.14	**400.08**
34	10	100	50	5	312.15	**297.31**
35	10	100	25	10	**403.31**	407.72
36	10	100	50	10	306.42	**305.12**
37	10	100	25	20	**201.63**	226.59
38	10	100	50	20	200.85	200.85
39	10	100	25	40	90.90	90.90
40	10	100	50	40	54.33	54.33
41	10	200	25	5	732.62	732.62
42	10	200	50	5	504.18	504.18
43	10	200	25	10	517.64	517.64
44	10	200	50	10	**490.06**	492.43
45	10	200	25	20	390.80	390.80
46	10	200	50	20	284.12	**283.56**
47	10	200	25	40	122.06	122.06
48	10	200	50	40	81.73	81.73

Table 3 details the results when there are 20 potential stops. According to the Table 3, the probability model is able to detect competitive sequences and many occasions getting less time than the genetic algorithm.

Table 3. Comparative results for 20 potential stops

id	stop	stud	cap	wd	GA	EDA
49	20	100	25	5	713.64	**659.51**
50	20	100	50	5	671.19	**601.85**
51	20	100	25	10	**535.17**	559.42
52	20	100	50	10	558.57	**494.50**
53	20	100	25	20	385.90	**381.34**
54	20	100	50	20	223.88	**220.32**
55	20	100	25	40	108.15	108.15
56	20	100	50	40	46.74	46.74
57	20	200	25	5	966.16	**938.74**
58	20	200	50	5	657.03	**609.33**
59	20	200	25	10	750.58	**724.65**
60	20	200	50	10	630.02	**615.54**
61	20	200	25	20	608.72	**559.40**
62	20	200	50	20	341.25	**318.49**
63	20	200	25	40	**153.75**	154.58
64	20	200	50	40	78.93	78.93
65	20	400	25	5	1329.41	1329.41
66	20	400	50	5	836.29	**807.82**
67	20	400	25	10	1007.80	1007.80
68	20	400	50	10	719.19	**672.55**
69	20	400	25	20	**918.31**	920.04
70	20	400	50	20	460.24	**425.94**
71	20	400	25	40	280.07	280.07
72	20	400	50	40	**118.33**	122.49

Table 4 depicts on the results when there are 40 potential stops. Based on the results shown in the Table 4, the probability model is able to detect competitive sequences and the most cases getting less time than the genetic algorithm.

Table 4. Comparative results for 40 potential stops

id	stop	stud	cap	wd	GA	EDA
73	40	200	25	5	1414.20	**1351.84**
74	40	200	50	5	1254.45	**1183.60**
75	40	200	25	10	1191.57	**1087.14**
76	40	200	50	10	916.95	**858.14**
77	40	200	25	20	584.50	**547.98**
78	40	200	50	20	429.42	**383.73**
79	40	200	25	40	**160.35**	163.88
80	40	200	50	40	111.30	111.30
81	40	400	25	5	1775.80	**1703.86**
82	40	400	50	5	1523.28	**1462.30**
83	40	400	25	10	1222.86	**1205.62**
84	40	400	50	10	1310.81	**1211.93**
85	40	400	25	20	957.39	**919.50**
86	40	400	50	20	726.32	**679.67**
87	40	400	25	40	299.85	299.85
88	40	400	50	40	**106.93**	107.14
89	40	800	25	5	2982.74	**2980.99**
90	40	800	50	5	1823.20	**1721.53**
91	40	800	25	10	2527.08	**2458.14**
92	40	800	50	10	1398.47	**1348.72**
93	40	800	25	20	1672.09	**1662.70**
94	40	800	50	20	1048.45	**1015.14**
95	40	800	25	40	451.63	451.63
96	40	800	50	40	342.94	**337.28**

Table 5 provides the results when there are 80 potential stops. From Table 5 can see how the probability model is able to detect competitive sequences and almost always getting less time than the genetic algorithm.

Table 5. Comparative results for 80 potential stops

id	stop	stud	cap	wd	GA	EDA
97	80	400	25	5	3183.28	**2971.76**
98	80	400	50	5	3206.93	**3037.08**
99	80	400	25	10	2308.92	**2215.67**
100	80	400	50	10	1751.58	**1728.20**
101	80	400	25	20	1228.60	**1180.88**
102	80	400	50	20	765.17	**753.68**
103	80	400	25	40	277.24	**275.95**
104	80	400	50	40	**164.84**	165.01
105	80	800	25	5	3792.85	**3708.18**
106	80	800	50	5	3119.84	**3033.81**
107	80	800	25	10	2777.51	**2713.75**
108	80	800	50	10	2225.75	**2158.87**
109	80	800	25	20	2037.22	**1964.91**
110	80	800	50	20	1292.76	**1183.44**
111	80	800	25	40	**480.24**	485.03
112	80	800	50	40	271.94	**271.77**

Figure 4 shows the dispersion and symmetry for both algorithms for each number of potential stops. As the dispersion and symmetry observed between the values obtained is very similar, so it follows that both algorithms are equally stable in finding better sequences (routes).
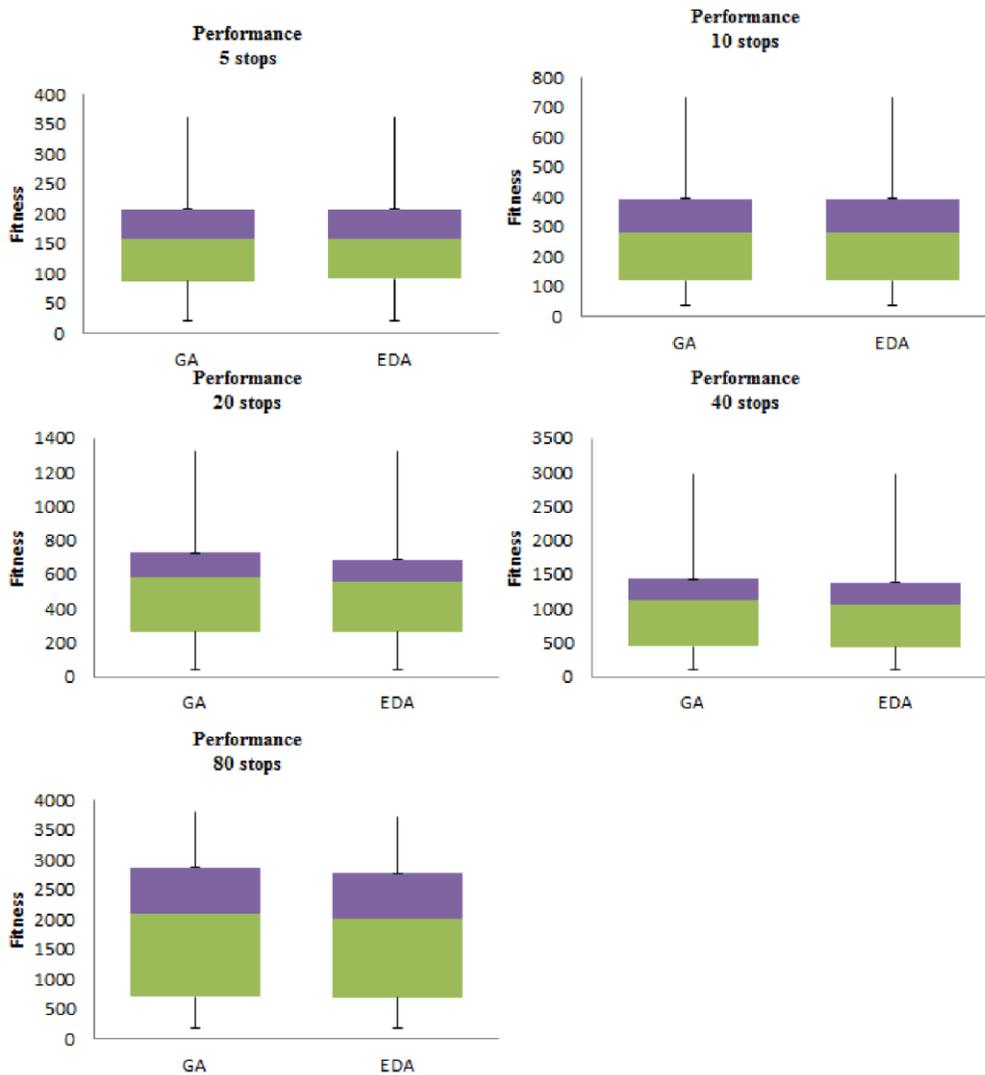
Fig. 4. Dispersion and symmetry between both algorithms.

Further, a hypothesis test is performed to determine whether the proposed EDA is better than the genetic algorithm. Table 6 describes when there is a statistically significant difference between the algorithms.

Table 6. Hypothesis testing between algorithms

Stops	t-test	t-critical	µ GA	µ EDA	σ GA	σ EDA
5	0.037	0.0630	158.16	159.17	94.43	93.34
10	0.036	0.0630	278.59	276.75	175.77	175.09
20	**0.199+**	0.0630	545.81	526.57	336.92	330.88
40	**0.191+**	0.0630	1093.02	1052.23	744.19	730.02
80	**0.155+**	0.0632	1805.29	1740.50	1205.10	1157.52

4 Conclusions

Based on the above results, the contribution of this research to the solution of permutation-based representation problems such as the school bus routing problem with stops selection is confirmed. The probability model used is able to find better solutions to reduce the travel time for a set of school buses. The proposed model makes explicit the search, a situation that is random in the variation process on other algorithms. The construction mechanism of the solutions permits to determine the chance for a bus stop to be chosen on the route through the probability model. The method of allocating student-stop used in this research was no downside to reduce travel time by the model used.

References

1. B. Minocha and S. Tripathi, "Solving School Bus Routing Problem Using Hybrid Genetic Algorithm: A Case Study," in Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), vol. 236, B. Babu, A. Nagar, K. Deep, M. Pant, J. Bansal, K. Ray, et al., Eds. Springer India, 2012, pp.93-103.
2. P. Schittekat, J. Kinable, K. Sörensen, M. Sevaux, F. Spieksma and J. Springael, "A metaheuristic for the school bus routing problem with bus stop selection," European Journal of Operational Research, vol. 229, pp. 518-528, 2013.
3. H. Niu, "Application of Genetic Algorithm to Optimize Transit Schedule under Time-Dependent Demand," in Computational Intelligence for Traffic and Mobility, vol. 8, W. Wang and G. Wets, Eds. Atlantis Press, 2013, pp. 71-88.
4. O. Díaz-Parra, J. Ruiz-Vanoye, M. Buenabad-Arias and A. Canepa-Saenz, "Vertical Transfer Algorithm for the School Bus Routing Problem," in Transactions on Computational Science XXI, vol. 8160, M. Gavrilova, C. Tan and A. Abraham, Eds. Springer Berlin Heidelberg, 2013, pp. 211-229.
5. J. Euchi, and R. Mraihi, "The urban bus routing problem in the Tunisian case by the hybrid artificial ant colony algorithm," Swarm and Evolutionary Computation, vol. 2, pp. 15-24, 2012.
6. J. Pacheco, R. Caballero, M. Laguna, and J. Molina, "Bi-Objective Bus Routing: An Application to School Buses in Rural Areas," Transportation Science, vol. 47(3), pp. 397-411, 2013.
7. J. Park, and B.I. Kim, "The school bus routing problem: A review," European Journal of Operational Research, vol. 202(2), pp. 311-319, 2010.
8. L. Wang, S. Wang, Y. Xu, G. Zhou and M. Liu, "A bi-population based estimation of distribution algorithm for the flexible job-shop scheduling problem," Computers and Industrial Engineering, vol. 62, pp. 917-926, 2012.
9. C. Prins, "A simple and effective evolutionary algorithm for the vehicle routing problem," Computers & Operations Research, vol. 31(12), pp. 1985-2002, 2004.