Sigma Journal of Engineering and Natural Sciences, Vol. 42, No. 6, pp. XX-XX, December, 2024
Yildiz Technical University Press, Istanbul, Turkiye

Two-stage clustering and routing problem by using FCM and K-means with Genetic Algorithm

Ebru PEKEL ÖZMEN[1,*], Tarık KÜÇÜKDENİZ[2]

Abstract

The School Bus Routing Problem (SBRP) is a challenging optimization problem that has received increasing attention in recent years. The problem is composed of three sub-problems: facility location selection, assignment problem, and vehicle routing problem, which can be solved in a single stage or across multiple stages. In this study, we propose a novel two-stage approach to solve the SBRP that combines Fuzzy C Means (FCM) and K-means clustering algorithms with a Genetic Algorithm (GA). In the first stage, we used FCM and K-means to identify the optimal bus stop locations and assigned students to the nearest stop based on the distance metric. This two-stage approach reduces the search space and improves the efficiency of the GA in the second stage. In the second stage, we employed the GA to generate the optimal vehicle route that minimizes the total distance traveled by all vehicles. We compared our results with those in the literature and found that the K Means-GA approach outperformed the previous results. However, the FCM-GA approach yielded significantly inferior results, indicating that the choice of clustering algorithm plays a crucial role in the performance of the overall system. Our study provides insights into the importance of selecting appropriate clustering algorithms for solving the SBRP and proposes a two-stage solution that can be easily implemented in real-world scenarios. Our approach reduces the computational time and provides an effective solution for reducing the total distance traveled by school buses.

Keywords: *LAR, SBRP, Genetic Algorithms, Fuzzy Clustering, K-means*

Introduction

With urbanization, urban transportation has become a major concern for many large cities. However, public transit systems are often perceived as inefficient, leading people to rely on their own vehicles or taxis instead. In the context of school transportation, the inefficiency of school-provided transportation systems can result in an increase in private vehicle usage, which contributes to traffic congestion, air pollution, and environmental harm. Therefore, it is essential to develop efficient and cost-effective transportation systems for students that prioritize their comfort and safety while reducing the overall environmental impact.

Several studies have addressed the school bus routing problem (SBRP) in recent years. These studies have proposed various mathematical models and metaheuristic algorithms to improve the efficiency of school transportation systems while balancing the service provider's costs and students' convenience. Some studies have considered factors such as students' travel patterns, class periods, and school bell times to optimize school bus routing. However, there is still room for improvement in terms of minimizing the walking distance of individuals and the total distance traveled by vehicles, especially in the context of the first cluster and second route approach. Miranda, de Camargo [1] have discussed SBRP by extending it with mixed loads. Oluwadare, Oguntuyi [2] proposed a Genetic Algorithm to solve the school bus route problem. Sun, Duan [3] addressed the school bus routing problem in a stochastic and time-dependent road environment.

Mahmoudzadeh and Wang [4] tried to identify a feasible technique for scheduling the university shuttle fleet, considering students' travel patterns and class periods. The annual ridership data for 2018–2019 of the Texas A&M University (TAMU) shuttle system were utilized to detect unique weekday and weekend travel patterns. A clustering strategy was presented to create uniform departure times within the current schedules for on-campus and off-campus routes, considering the high student loads and graduate/undergraduate habits. [5] studied a case for the capacitated employee bus routing problem using data from a retail company. They constructed a mathematical model to minimize the total distance of bus routes. The number and location of bus stops was identified using k-means and fuzzy c-means clustering techniques. Then, due to the NP-Hard nature of the bus routing problem, simulated annealing (SA) and genetic algorithm (GA)-based solutions were presented. A hybrid

This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkilic
[1] Department of Industrial Engineering, Samsun University, Samsun, Türkiye
[2] Department of Industrial Engineering, Istanbul University-Cerrahpaşa, Istanbul, Türkiye
** E-mail address: ebru.pekel@samsun.edu.tr*
Orcid id: https://orcid.org/0000-0001-7717-6790 Ebru Pekel Özmen, 0000-0002-6670-1809 Tarık Küçükdeniz
Manuscript Received 27 January 2023, Revised 12 March 2023, Accepted 04 May 2023

iterative local search (ILS) metaheuristic algorithm for SBRP with various planning scenarios, such as homogeneous or heterogeneous fleets, single load or mixed load modes of operation, is an attempt made in [6]. Pérez, Ansola [7] created a mathematical model based on literature and modified it such that metaheuristics could be used to solve it.

Incorporating school bell times as decision variables rather than input data allowed them to extend the traditional school bus routing problem [8]. The authors used the Memetic algorithm to solve the proposed model and argue that changing school working hours (bell ringing times) is especially helpful in reducing transportation costs. Calvete, Galé [9] consider the SBRP if students are free to select the bus stop of their choice. It is created and presented to convert a two-stage optimization model to a single-level mixed integer linear programming (MILP) model. To address the issue, a straightforward and efficient metaheuristic algorithm has been created.

Pérez, Sánchez-Ansola [10] introduced a novel method for reducing execution time, which is one of the factors considered while solving a problem using SBRP. Hou, Liu [11] proposed a local search-based metaheuristic algorithm framework for SBRP based on the analysis of SBRP problem models. The proposed framework provides metaheuristic development for the Capacitated Vehicle Routing Problem by reducing or modifying the constraints of SBRP. Experimental results of several examples show that the metaheuristic algorithm based on the proposed approach can be implemented quickly and applied to different SBRP applications.

Feng, Zhang [12] proposed two strategies (mixed-load and single-load) for the mixed-load school bus routing problem (SBRP) to improve student commuting efficiency while reducing operational costs. A bi-level programming model was constructed to maximize school accessibility and minimize the number of school buses. The entropy-based weight method was used to evaluate the strategies, and case studies were conducted using school bus data in Dalian, China. [13] addressed SBRP with student input in bus stop selection. A bilevel optimization model with multiple followers was proposed and transformed into a single-level mixed integer linear programming model. A metaheuristic algorithm was developed to solve the problem, which involved solving four MILP problems to obtain upper bounds. Computational experiments using benchmark instances showed the algorithm was effective in terms of solution quality and required computing time.

In this study, the School Bus Routing Problem (SBRP) was addressed in a two-stage approach. Initially, the Fuzzy C Means (FCM) method was utilized to identify the optimal locations for bus stops. Subsequently, students were assigned to the nearest bus stop. Alongside FCM, the K-means algorithm was also employed to pinpoint the optimal bus stop locations. In the second stage, the Genetic Algorithm (GA) was employed to generate the most efficient vehicle route with the goal of minimizing the total distance traveled by all vehicles.

The remainder of this paper is organized as follows: In the second section, the literature review is presented. In the third section, the proposed approaches are introduced. The fourth part shows experimental results to compare the performances of the proposed approaches. Finally, the discussion and conclusions are presented.

Material and Methods

This section presents the materials and methods used in this study to solve the SBRP problem. The first part introduces the K Means algorithm used to identify the bus stop locations, while the second part explains the Fuzzy C-means (FCM) algorithm used as an alternative method. Additionally, the Genetic Algorithm (GA) is described to be utilized in the second stage.

K-Means Algorithm

The K-means method is an iterative technique that attempts to divide the dataset into K unique, non-overlapping subgroups (clusters), where each data point belongs to only one group. It attempts to make intra-cluster data points as comparable as feasible while maintaining clusters as distinct as possible. It allocates data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (the arithmetic mean of all data points belonging to that cluster) is the smallest possible value. Data points within the same cluster are more homogenous (similar) the less variance there is within clusters. The implementation of the K-means algorithm is as follows ([14]:

Step 1: Indicate the number of cluster K.

Step 2: Initialize centroids by first shuffling the dataset and then picking K data points at random without replacement for use as centroids.

Step 3: Repeat iterating until the centroids do not change. Thus, the assignment of data points to clusters remains unchanged.

Fuzzy C Means (FCM)

The fuzzy C-Means method is a variant of the K-Means algorithm that employs fuzzy logic and seeks to find the least value of the objective function. It is possible for a data sample to belong to more than one cluster at the same time in FCM. Similarity is indicated by the membership value. In FCM, a membership value and a data sample are assigned based on cluster center and similarity ([15]. Membership values range from 0 to 1, and the higher the similarity, the higher the membership value. Steps of the Fuzzy C-Means algorithm are given in Equations 1-3 ([15].

$$O_{\mu}(M,y) = \sum_{i=1}^{c} \sum_{k=1}^{N} (m_{ik}^{\mu}) \|x_i - y_k\|^2 \tag{1}$$

where y_j is the mean for that points over cluster i, $\|x_i - y_k\|^2$ is a squared inner product distance norm, and m_{ik}^{μ} represents the fuzzy partitions, where m_{ik} denotes the membership degree that the ith data point belongs to the kth cluster and μ is a scalar for controlling the fuzziness of the resulting clusters. Besides, x represents the geographical X and Y coordinates.

Step 1: Assign membership matrix $M^{(0)} = [m_{ik}]$.

Step 2: Compute cluster centers using Equation 2.

$$y_k = \frac{\sum_{i=1}^{n} (m_{ik})^{\mu} x_i}{\sum_{i=1}^{n} (m_{ik})^{\mu}} \tag{2}$$

Step 3: Arrange membership values using Equation 3.

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\left| \frac{x_i - y_k}{x_i - c_j} \right| \right)^{\frac{2}{\mu - 1}}} \tag{3}$$

Step 4: If the differences in total costs estimated by consecutive membership matrices are less than the stopping criteria, the algorithm terminates.

In those equations, k and N represent the number of clusters and students, respectively. x_i means the coordinates of the student i, and y_k means the coordinates of the cluster k.

Genetic Algorithm (GA)

Genetic algorithms are stochastic algorithms whose search techniques are based on biological genetics and natural evolution concepts. Holland first presented the fundamental ideas of genetic algorithms in 1975. Individuals in a population are represented as chromosomes in this methodology, and a sequence of genetic operations is performed. It is widely acknowledged that the possible solution to each problem is an individual who may be represented using a set of parameters. Every cell in nature contains chromosomes that make up the DNA of that cell. Chromosomes contain the genes that regulate a living organism's character differentiation. An individual or chromosome is the potential solution to a problem in a genetic algorithm. The population is made up of individuals. Population, on the other hand, refers to the sum of the genetic algorithm's solutions. Mating, mutation, and reproduction are all genetic processes that are employed in genetic algorithms. Genetic algorithms offer an ideal setting for tackling large-scale optimization challenges. Although it belongs to the class of probabilistic algorithms, it differs from random number generation methods. Genetic operators, such as mutation, can readily escape local optimal points in genetic algorithms ([16]. The details of the method are not provided in this paper because it is a well-known universal algorithm.

A typical GA focuses on two things: fitness function and solutions. Constraints cannot be defined as in linear programming ([17]. As a result, the solution space and fitness function must be suitably adjusted. The pseudocode of GA to solve the proposed SBRP is presented in Table 1.

Table 1. Pseudo code of the applied metaheuristic algorithm for proposed SBRP

Stage 0:
Obtain student's coordinates P_i
Cap: capacity of the vehicle
Stage 1:
Find the cluster center with FCM
Assign to C_x and C_y
Stage 2:
$x_{values} \leftarrow random(0{,}10) * n$
Generate x_{values} by GA
Stage 3:
Calculate distance between the $distance_k = \{P_i, C_i\}$
$\sum_{student} walking\ distance\ for\ each\ student + \sum_{arc} total\ travelled\ distance$

Cluster-First Route-Second Approach for School Bus Routing Problem

Mathematical Model for multi-vehicle SBRP

The aim of the mathematical model is to design a system that will minimize the walking distance of each student to the assigned stops and minimize the total distance traveled by the vehicles. The mathematical model of the proposed SBRP problem is a bi-objective mixed integer location routing model. Even the single-objective location routing problem is an NP-hard problem because it combines the two problems, which are the facility location and vehicle routing problem and the location routing problem ([10]. Because the bi-objective position routing issue is more complicated, it is also NP-hard. The representation of the model, its parameters, and the decision variables are presented below.

Model A:

Table 2. Symbols used in the mathematical model

Indices	
m	indices of the student *(m= 1,2,…M)*
i, j	*indices of the* stops *(i=1,2,3…*C_i)
v	*vehicle*
Parameters	
C_i	$the\ number\ of\ the\ stop$
P_{mx}	$coordinate\ value\ of\ the\ m^{th}\ student\ on\ the\ x-axis$
P_{my}	$coordinate\ value\ of\ the\ m^{th}\ student\ on\ the\ y-axis$
W	$the\ capacity\ of\ the\ vehicle$
Models Variables	
Z_{iv}	$\{if\ the\ vehicle\ v\ visits\ to\ i^{th}\ stop, \quad 1\ otherwise, \quad 0$
Y_{im}	$\{if\ m^{th}\ student\ assigns\ to\ i^{th}\ stop, \quad 1\ otherwise, \quad 0$
d'_{im}	$the\ distance\ between\ i^{th}\ stop\ and\ m^{th}\ student$

d''_{ij}	$the\ distance\ between\ i^{th}\ stop\ and\ j^{th}\ stop$
Decision Variables	
C_{ix}	$Coordinate\ value\ of\ the\ \ i^{th}\ stop\ on\ the\ x-axis$
C_{iy}	$Coordinate\ value\ of\ the\ \ i^{th}\ stop\ on\ the\ y-axis$
x_{ijv}	$\{if\ the\ vehicle\ v\ travels\ from\ i^{th}\ stop\ to\ j^{th}\ stop, \quad 1\ \ otherwise, \quad 0$

Model:

$$z = \sum_m \sum_i d'_{im} * Y_{im} + \sum_i \sum_j \sum_v d''_{ij} * x_{ijv} \tag{4}$$

$$d'_{im} = |P_{mx} - C_{ix}| + |P_{my} - C_{iy}| \qquad \forall\, m, i \tag{5}$$

$$\sum_i Y_{im} = 1 \qquad \forall\, m \tag{6}$$

$$d''_{ij} = \sqrt{\left(C_{ix} - C_{jx}\right)^2 + \left(C_{iy} - C_{jy}\right)^2} \qquad \forall\, i, j \tag{7}$$

$$\sum_i \sum_v x_{ijv} = 1 \qquad \forall\, j \tag{8}$$

$$\sum_j \sum_v x_{ijv} = 1 \qquad \forall i \tag{9}$$

$$\sum_{i=0, j=0}^{Ck} x_{ijv} \left(\sum_m Y_{im} \right) = W \qquad \forall\, v \tag{10}$$

$$\sum_i x_{ijv} = Z_{jv} \qquad \forall\, j, v \tag{11}$$

$$\sum_j x_{ijv} = Z_{iv} \qquad \forall\, i, v \tag{12}$$

$$\sum_i x_{ijv} = \sum_h x_{jhv} \qquad \forall\, i, v \tag{13}$$

$$\sum_i x_{0iv} = \sum_j x_{j0v} = 1 \qquad \forall\, v \tag{14}$$

Equation 4 expresses the objective function. It minimizes the walking distance of students and the total distance traveled by vehicles. Equation 5 and Equation 7 calculate the walking distance of students and the distance between cluster centers, respectively. Equation 6 checks that each student is assigned to only one cluster. Equations 8–9 state that only one vehicle should be assigned between any two points. Equation 10 verifies that the number of students served by the vehicle along the entire route of each vehicle does not exceed the vehicle's capacity. Equations 11–12 state that a bus arriving at a bus stop also leaves the station and stops only on its route. Equation 11 ensures that the flow of each vehicle is maintained. Equation 14 states that each bus departs from school and returns to school. Equations 15–16 are MTZ equations.

Miller-Tucker-Zemlin (MTZ) formulation uses an extra positive variable called as p_i. p_i takes a value for each node except the school. If a vehicle is traveling from node i to node j, p_j must be greater than p_i. The value of p_i grows with each subsequent call point.

$$p_i - p_j - n * \sum_{v}^{V} x_{ijv} \leq n - 1 \qquad 2 \leq i \neq j \leq n \tag{15}$$

$$0 \leq u_i \leq n - 1 \qquad 2 \leq i \leq n \tag{16}$$

The process of selecting bus stops and generating bus routes is frequently referred to as a location-routing problem (LRP), which is an example of an NP-hard problem. An NP-hard problem is one for which there are no known polynomial-time algorithms that can provide an optimal solution for every instance of the problem. In such situations, dividing the problem into parts helps facilitate its resolution.

The foundational "divide and conquer" approach served as the primary inspiration for the creation of this heuristic, which was built on that method. An assignment model is first solved to establish a cluster of pickup locations for each vehicle using the cluster-first and route-second approaches. Subsequently, the traveling salesman problem (TSP) is addressed for all clusters. Finally, a route is generated using the TSP solutions.

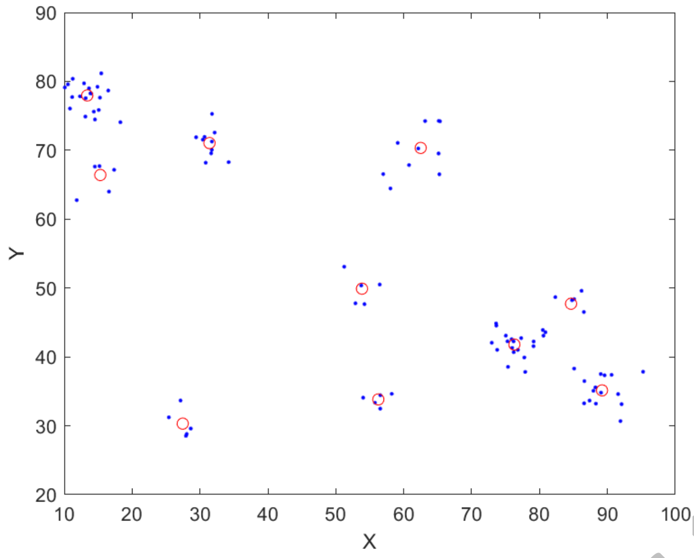

Figure 1. Sample cluster centers for n=10

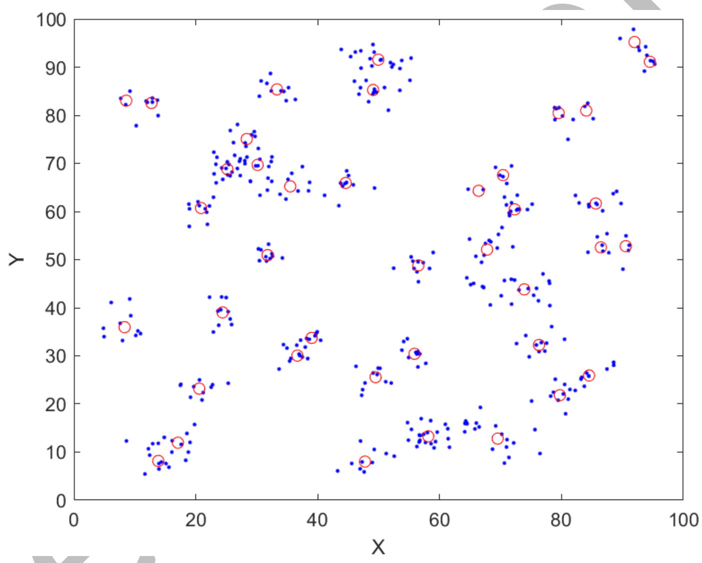

Figure 2. Sample cluster centers for n=40

The results obtained from the fuzzy C-means algorithm under different conditions are presented in Figures 1 and 2. Figure 1 depicts the cluster centers obtained for a scenario involving 10 bus stops, whereas Figure 2 presents the cluster centers for a scenario with n = 40 bus stops. It is common practice to locate bus stops in parts of a community that have a high concentration of student inhabitants. The decision variables C_{ix} and C_{iy} in Model A may be derived from the outputs that are provided by the FCM. The last variable that must be determined is the one that is referred to as X_{ijv}.

Route Generation

The most expedient route is determined with the help of GA. An initial solution is chosen at random by the genetic algorithm. The resolution that was randomly given is called an initial chromosome. If we assume that the value of each gene (r_i) falls somewhere between 0 and 1, then the composition of a chromosome with 5 stops is as follows:

$$X_{ijv} = (r_1, r_2, \dots, r_5)$$

Each gene value is sorted from lowest to highest, and the final order of the rankings is kept. This rank indicates the order in which the bus stops are located along the route. When the number of students on the bus that is making the stops exceeds the capacity of the vehicle, another bus is brought into service. There is a limit of one stop per bus when overbooking is permitted.

Table 3. Route Generetion by GA

1. Generate the initial solution
2. Assign the vehicles to each bus stop
3. Compute the objective function
   3.1. Modify the solution by genetic operators
   3.2. Compute the objective function
4. Kept the best solution

The calculations are done based on the premise that the school serves as both the beginning and the finish point for the vehicle routes. The fundamental procedures that GA employs to determine the most efficient route are summarized in Table 3.

Experimental Results and Discussion

The dataset originally consisted of 112 cases, all of which were taken from [18]. Sample sizes vary anywhere from 5 stops and 25 students to 80 stops and 800 students. In addition, the dataset took into consideration a maximum of four different maximum walking distances. Since the maximum walking distance was not included as a parameter in this study, there were no experiments conducted to investigate the effects of varying the maximum walking distance. As a result, the experiments were only conducted in a total of 28 instances. The instances have a standard fleet of buses that can accommodate 25 or 50 students.

The parameter values of the employed metaheuristic algorithm are shown in Table 4. The maximum value of the decision variable is set to 1 in GA, while the minimum value is set to 0.

Table 4. Parameters of GA

Parameters	Values
Max_num_iteration	1500
Population_size	100
Mutation_rate	0.1
Elit_ratio	0.01
Crossover_rate	0.5
Parents_portion	0.3
Crossover_type	uniform
Variable_bounds	(0,1)

Table 5 contains the results for the problem instances. The number of stops is indicated in the stop column. The *Student* column specifies the number of students served, while the *CAP* column indicates the bus capacity. The students are transported by a homogeneous fleet of vehicles of identical capacity. Total walking distance is shown by the total walking column, while total driving distance is indicated by the total driving distance.

The *Total driving** columns are vital to compare *former_best* and *new_best* value. As an example, the best value is found as 241.68 for Instance 09, which is found in this work; also, this value is obtained as 286.68 in [18] as seen in Table 5. It suggests that the proposed approach improves performance by 15.7%.

The purpose of this research was not only to lower the total distance traveled by vehicles but also the total distance traveled by pedestrians. When the first phase of the clustering procedure is complete, it is verified that the students' stopping locations are placed at the geographically closest points to them. This reduces the amount of walking distance these students must travel. In addition, the predefined bus stop locations were used to find the optimal routes.

Table 5. Results for FCM-GA

				FCM & GA		
ID	#Stop	#Student	CAP	Total walking	Total driving*	Former_Best
1	5	25	25	159.26	266.26	141.01
2	5	25	50	151.88	250.60	161.62
9	5	50	25	254.00	241.68	286.68
10	5	50	50	254.00	229.09	197.20
17	5	100	25	724.46	253.40	360.35
18	5	100	50	600.96	256.95	304.23

25	10	50	25	318.48	251.60	282.12
26	10	50	50	181.51	300.15	244.54
33	10	100	25	464.19	482.80	403.18
34	10	100	50	319.94	286.28	296.53
41	10	200	25	970.05	468.94	735.27
42	10	200	50	983.11	367.40	512.16
49	20	100	25	354.82	478.61	520.24
50	20	100	50	338.29	409.17	420.64
57	20	200	25	409.17	647.29	903.84
58	20	200	50	734.19	489.53	485.65
65	20	400	25	1937.84	738.64	1323.35
66	20	400	50	1937.84	594.74	733.54
73	40	200	25	538.39	794.48	831.94
74	40	200	50	631.57	771.73	593.35
81	40	400	25	1292.30	1185.69	1407.05
82	40	400	50	1333.54	829.79	858.80
89	40	800	25	2607.60	1584.66	2900.14
90	40	800	50	2736.71	1054.22	1345.70
97	80	400	25	1196.71	2571.42	1546.23
98	80	400	50	1038.53	2052.97	1048.56
105	80	800	25	2139.61	2734.03	2527.96
106	80	800	50	2405.05	2376.25	1530.58

Table 5 displays the results of the suggested approach, indicating that favorable performance was achieved in 17 out of the 28 analyses. It is noteworthy that the "driving" column is of particular significance for the purpose of comparison, while the "walking" column represents an additional accomplishment that we aim to achieve. This is because students are required to walk less concurrently, which has had a notable impact on the identification of bus stop locations. Consequently, the search for the shortest driving distance is complicated by the need to consider the walking distance factor. Despite the multiplicity of focus points, the suggested approach still yielded better results in more than half of the instances.

Table 6 demonstrates the results of the K-means and GA methodology, indicating that this strategy outperformed the suggested approach in 19 of the 28 evaluations. However, it is important to note that as the sample size grows, the two-stage strategy's performance has shown signs of deterioration.

Table 6. Results for K means-GA

K-means & GA						
ID	#STOP	#STUDENT	CAP	Total walking	Total driving*	Former_Best
1	5	25	25	137.57	146.05	141.01
2	5	25	50	276.78	157.07	161.62
9	5	50	25	542.70	227.65	286.68
10	5	50	50	387.76	182.66	197.20
17	5	100	25	1096.30	262.52	360.35
18	5	100	50	534.27	253.06	304.23
25	10	50	25	421.21	225.63	282.12

26	10	50	50	301.48	281.13	244.54
33	10	100	25	429.43	280.73	403.18
34	10	100	50	433.37	271.95	296.53
41	10	200	25	1278.26	422.71	735.27
42	10	200	50	1570.33	375.02	512.16
49	20	100	25	415.70	463.32	520.24
50	20	100	50	434.92	417.95	420.64
57	20	200	25	1266.32	625.01	903.84
58	20	200	50	740.33	473.40	485.65
65	20	400	25	1829.96	740.13	1323.35
66	20	400	50	1723.89	499.52	733.54
73	40	200	25	853.92	906.26	831.94
74	40	200	50	600.75	702.89	593.35
81	40	400	25	1198.97	1084.13	1407.05
82	40	400	50	1463.16	926.50	858.80
89	40	800	25	3065.63	1612.96	2900.14
90	40	800	50	2810.92	1183.00	1345.70
97	80	400	25	1197.36	2485.06	1546.23
98	80	400	50	1059.77	2440.06	1048.56
105	80	800	25	2244.93	2876.94	2527.96
106	80	800	50	2488.89	2360.54	1530.58

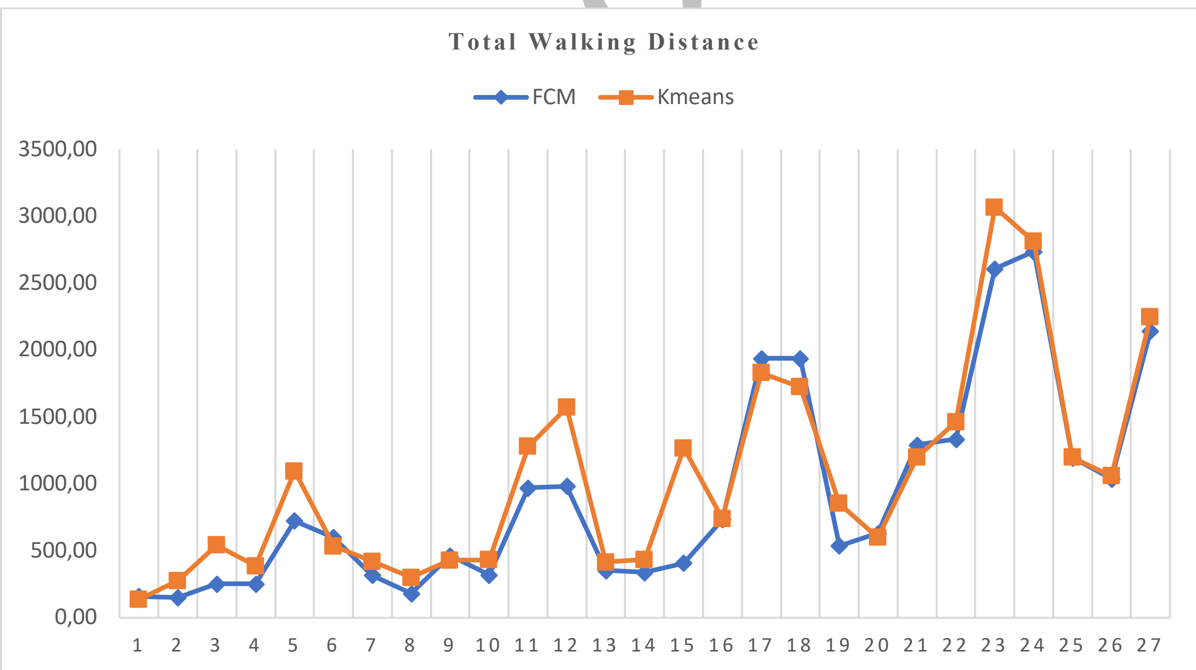

Figure 3. Comparison of the total walking distances

This study aimed to compare the FCM and K-means clustering algorithms, as depicted in Figure 3. The graph illustrates the results obtained from each algorithm, and a detailed analysis of the graph revealed that the K-means approach demonstrated superior efficiency in reducing the walking distance for students.

Furthermore, Table 7 provides a comprehensive statistical summary of the proposed approaches. The K-Means-GA approach yielded the best mean value of 817.28, surpassing the previous results reported in the literature which stood at 817.945. However, given that the mean alone may not suffice for comparison purposes, the median value was also considered. Remarkably, the K-Means-GA approach also resulted in the smallest median value.

In summary, the findings of this study provide evidence that the K-means clustering algorithm, particularly the K-Means-GA approach, is a promising method for optimizing bus stop locations and reducing walking distances for students.

Table 7. Statistical summary of the proposed approaches

	FCM-GA	K-Means-GA	Former_Best
Mean	820.2983	817.2809	817.945
Median	486.1672	468.3602	556.795
Min	229.0923	146.0465	141.01
Max	2734.026	2876.941	2900.14

Discussion and Conclusion

This paper presents a two-stage approach to the school bus routing problem, with the aim of reducing both the total distance traveled by vehicles and students. The first stage involves selecting suitable bus stops using the FCM and K-Means algorithms, while the second stage uses GA to plan bus routes. The dataset utilized in this study was obtained from an easily accessible source. The application process was conducted in a total of 28 distinct instances, wherein both the Fuzzy C Means and the GA implementations were employed to attain optimal outcomes. Analysis of the experimental results revealed that the proposed approach yielded significant advantages overall.

The proposed approach offers decision makers possible solutions and trade-offs between the objectives. The study shows that the proposed approach offers significant benefits, including a considerable reduction in solution time. However, there are some limitations, such as the lack of coordination across phases' objective functions.

A brief summary of the fundamental points outlined in this study is presented below.:

- The proposed two-stage approach reduces solution time and enables modifications at any stage.
- Overbooking is a common occurrence in real-life scenarios, but it is rarely discussed in the SBRP literature.
- The unchangeable rule that prohibits exceeding the maximum capacity of a vehicle's seating capacity has been loosened in this study.
- The first bus stop locations are uncertain, and there is no available sequence of potential bus stops.

In future studies, it may be beneficial to investigate a method that integrates both objective functions into a single phase, potentially enhancing the efficiency and effectiveness of the solution. However, it is important to note that the utilization of a goal programming approach in a one-step process may result in an increased level of complexity in the problem formulation and solution procedure. Thus, the aim of any such approach should be to provide reasonable and convincing results in a practical and feasible manner.

REFERENCES

[1]. Miranda, D.M., et al., *A multi-loading school bus routing problem.* Expert Systems with Applications, 2018. **101**: p. 228-242.

[2]. Oluwadare, S.A., I.P. Oguntuyi, and J.C. Nwaiwu, *Solving school bus routing problem using genetic algorithm-based model.* International Journal of Intelligent Systems and Applications, 2018. **12**(3): p. 50.

[3]. Sun, S., Z. Duan, and Q. Xu, *School bus routing problem in the stochastic and time-dependent transportation network.* PloS one, 2018. **13**(8): p. e0202618.

[4]. Mahmoudzadeh, A. and X.B. Wang, *Cluster Based Methodology for Scheduling a University Shuttle System.* Transportation Research Record, 2020. **2674**(1): p. 236-248.

[5]. Kiriş, S.B. and T. Özcan, *Metaheuristics approaches to solve the employee bus routing problem with clustering-based bus stop selection*, in *Artificial Intelligence and Machine Learning Applications in Civil, Mechanical, and Industrial Engineering*. 2020, IGI Global. p. 217-239.

[6]. Xie, Y., et al., *A Metaheuristic with Learning Mechanism for Solving the Multi-school Heterogeneous School Bus Routing Problem.* IAENG International Journal of Computer Science, 2021. **48**(4).

[7]. Pérez, A.C.P., E.S. Ansola, and A.R. Suárez, *A Metaheuristic Solution for the School Bus Routing Problem with Homogeneous Fleet and Bus Stop Selection.* Ingeniería, 2021. **26**(2): p. 233-253.

[8]. Miranda, D.M., et al., *A metaheuristic for the rural school bus routing problem with bell adjustment.* Expert Systems with Applications, 2021. **180**: p. 115086.

[9]. Calvete, H.I., et al., *The school bus routing problem with student choice: a bilevel approach and a simple and effective metaheuristic.* International Transactions in Operational Research, 2021.

[10]. Pérez, A.C., et al., *A partial evaluation approach for the School Bus Routing Problem.* Heliyon, 2022: p. e09291.

[11]. Hou, Y., et al., *A Local Search-based Metaheuristic Algorithm Framework for the School Bus Routing Problem.* Engineering Letters, 2022. **30**(1).

[12]. Feng, R., et al., *School accessibility evaluation under mixed-load school bus routing problem strategies.* Transport Policy, 2023. **131**: p. 75-86.

[13]. Calvete, H.I., et al., *The school bus routing problem with student choice: a bilevel approach and a simple and effective metaheuristic.* International Transactions in Operational Research, 2023. **30**(2): p. 1092-1119.

[14]. Sinaga, K.P. and M.-S. Yang, *Unsupervised K-means clustering algorithm.* IEEE access, 2020. **8**: p. 80716-80727.

[15]. Küçükdeniz, T. and Ö. Erkal Sönmez. *Integrated Warehouse Layout Planning with Fuzzy C-Means Clustering*. in *International Conference on Intelligent and Fuzzy Systems*. 2022. Springer.

[16]. Pekel Özmen, E. and T. Özcan, *Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm.* Journal of Forecasting, 2020. **39**(4): p. 661-670.

[17]. Abubakar, A.M., et al., *Simplex C++ syntax for solving chemical engineering cost optimization problems.* Research Inventy: International Journal of Engineering And Science, 2021. **11**(7): p. 39-47.

[18]. Schittekat, P., et al., *A metaheuristic for the school bus routing problem with bus stop selection.* European Journal of Operational Research, 2013. **229**(2): p. 518-528.