

INTERNET OF THINGS-GROUP 4

PHASE-2

PUBLIC TRANSPORTATION OPTIMIZATION

STUDENT NAME : B.MOHAMMED HAPIPU

REGISTER NUMBER : 822621106302

NM ID : au822621106302

COLLEGE CODE : 8226

COLLEGE NAME : ARIFA INSTITUTE OF TECNOLOGY

EMAIL ID : mohammedhapipu@gmail.com

2. IoT Technology and Optimization

Algorithm for Vehicle Transportation Path

2.1. Key Technologies of the IoT

2.1.1. GPS Positioning and Wireless Sensor Network. There are 24 GPS satellites evenly distributed over the surface [8], and navigation information can be stored on the satellites [9].

The sensor node performs a preliminary fusion of the detected data and sends the data and its own position information to the sink node in the form of multihop transmission, and finally the detected data are sent to the end user via satellite and Internet.

The positioning method of the wireless sensor network is to equip a very small number of nodes with GPS, relying on the information interaction between the nodes, and using the positioning algorithm to calculate the location. This positioning process includes two stages: distance estimation and coordinate calculation. The calculation of the algorithm performance generally starts from the normal position and the weight of the issue. Among them, the calculation method of positioning accuracy is shown in the following formula. (1)

$$\text{Error } \sqrt{(x_i - p_i)^2 + (y_i - q_i)^2} \quad (1)$$

r

2.1.2. Data Fusion Method. Kalman filtering is an unbiased minimum method. The statistical properties of the measurement model are used to recursively determine the best fusion data estimate in the statistical sense. The Kalman filter uses the state space of a linear random system. In operation, the Kalman filter method uses the minimum variance of the linear deviation as the standard to change the state parameters of the filter [10].

Bayesian estimation is used for multiple sensor information fusion. We use the conditional probabilistic formula for processing and finally give a systematic decision based on several rules. Bayesian estimation is an effective method to fuse information from multiple sensors in a static environment.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (2)$$

Fuzzy theory is based on the basis of fuzzy sets, and the uncertain fuzzy number can be recognized by the computer after transformation. The received information is processed by fuzzy logic, and a rule basis is established based on experience. Finally, we compare the current state of sensor monitoring with the basic state to determine the

reliability of the information. If the information reliability is high, we use fuzzy variables to represent the actual variables and obtain the actual variable values by solving the fuzzy variables [11]. This method uses a systematic method to model the uncertainty of the fusion process and finally produces consistent fuzzy inferences. This approach can solve the problem of information conflict and inconsistency to some extent, but it requires the work of a confusing member, so the ability to learn and adapt is not enough.

2.1.3. Cloud Computing and Data Mining. Cloud computing is a new technology. At present, some large Internet companies have established super-servers to provide storage and computing data centers. The cloud database brings great convenience to small- and medium-sized enterprises and individual users. Cloud computing is an Internet-based computing model that the public participates in. Its computing resources (computing functions, storage functions, and interactive functions) are dynamic, scalable, and virtualized and can be provided as services [12].

The foundation of data mining is data warehouse. Data mining is the use of related mining tools to classify potential connections between data according to the needs of database information, thereby promoting the popularization and utilization of information. Finding understanding is a product of an organic combination of artificial intelligence and database, which enables it to understand native languages and have the ability to speak.

2.2. Optimization of Vehicle Transportation Path

2.2.1. Elements of the Problem of Vehicle Transportation Routes. The vehicle routing problem refers to the use of vehicles to provide delivery services to customers with various needs, optimize the overall distribution channels of the organization, maximize the satisfaction of the needs of each customer, and achieve the corresponding goals under the constraints of time and load [13]. The constituent elements of the vehicle routing problem include distribution centers, customers, vehicles, distribution network, parking lots, constraints, and goals [11]. Commonly used constraints include vehicle load constraints, travel distance constraints, and time window constraints. The usual goals to be achieved include the lowest total cost of delivery, the shortest delivery distance, and the shortest delivery time.

2.2.2. Mathematical Model of Vehicle Transportation Route Problem. The general transportation problem of automobiles is to make the total transportation distance the shortest or the transportation cost the lowest [14].

Define graph $G = (S, L)$, $S = \{s_1, s_2, \dots, s_n\}$, distribution route $L = \{(s_i, s_j), s_i, s_j \in S\}$,

distribution center s_0 , number of vehicles M , vehicle load G of each vehicle, customer

point $s_i (i = 1, 2, \dots, n)$, customer point demand (g_1, g_2, \dots, g_n) , and cost f_{ij} . The mathematical model established is as follows:

$$x_{ijm} = \begin{cases} 1, & \text{vehicle } m \text{ travels from point } i \text{ to point } j, \\ 0, & \text{otherwise,} \end{cases} \quad \begin{matrix} i, j = 0, 1, \dots, n; \\ m = 1, 2, \dots, M \end{matrix} \quad (3)$$

$$x_{im} = \begin{cases} 1, & \text{customer point } i \text{ is served} \\ & \text{by vehicle } m, \\ 0, & \text{otherwise,} \end{cases} \quad \begin{matrix} i, j = 0, 1, \dots, n; \\ m = 1, 2, \dots, M \end{matrix} \quad (4)$$

Assuming that the delivery vehicles are uniform and the delivery task is a one-way task, only the delivery problem needs to be considered. Formulas (3) and (4) are two decision variables. The goal is the lowest total transportation cost, and the optimization objective function is shown in the following formula:

$$\min Z = \sum_{i,j,m} f_{ij} x_{ijm}, \quad (5)$$

$$\sum_i g_i y_{im} \leq G, \quad (m = 1, 2, \dots, M), \quad (6)$$

$$y_{im} = 1, \quad (i = 1, 2, \dots, n), \quad m = 1, 2, \dots, M \quad (7)$$

$$\sum_m y_{im} = 1, \quad (i = 1, 2, \dots, n) \quad (8)$$

$$x_{ijm} = y_{mj}, \quad (j = 1, 2, \dots, n), \quad m = 1, 2, \dots, M \quad (9)$$

$$x_{ijm} = y_{mi}, \quad (i = 1, 2, \dots, n), \quad m = 1, 2, \dots, M \quad (10)$$

Among them, formula (6) indicates that the cargo loaded by the vehicle cannot exceed the load of the vehicle. Formula (7) indicates that each customer point is on the delivery route. Formula (8) indicates that the distribution center of each vehicle is consistent. Formulas (9) and (10) indicate that if the customer points i, j need to be on the delivery route of the vehicle m , the service of the vehicle can be obtained [15].

2.2.3. Classification of Vehicle Transportation Path Problems. The important feature of the static vehicle routing problem is that the variables related to the vehicle are known before the vehicle scheduling is arranged and will not change after the vehicle scheduling [16].

According to the number of optimized objective functions, it is divided into single-objective problem and multiobjective problem. According to the demand

characteristics, it can be divided into pure delivery, pure pickup, and integrated delivery and pickup vehicle routing problems. According to the vehicle type, it can be divided into single vehicle type and multiple vehicle type vehicle routing problems [17].

2.3. Intelligent Algorithm for Path Optimization

2.3.1. GA (Genetic Algorithm). GA can effectively solve the optimization problem of vehicle path. Even if the problem has discontinuities and nonlinearities or the internal structure of the problem is not clear, GAs can also solve these complex problems that cannot be mathematically modeled. The effective multibranch random parallel search feature makes the GA not easy to fall into the local optimum [18]. By increasing the number of populations, the optimization branches can be increased, which is very convenient for handling large-scale optimization problems, and the speed will also be faster.

First, change the optimal variables of the optimization problem to be optimized, then establish a function for individual compatibility calculation based on the function objectivity of the optimization problem. And randomly develop a group of individuals as the first solution point. If the conditions are not met, copy, cross, and mutate operations will be performed to verify again whether the termination conditions are met. Repeat the operation until the optimal solution of the degree of adaptation is obtained [19].

The fitness function is the optimization decision variable of the problem, which needs to be directly converted from the objective function of the mathematical model of the optimization problem. Its expression is shown in the following formula:

$$\zeta(x) = \text{Fit}(f(x)). \quad (11)$$

Considering that the probability requirement in the operation of GA is nonnegative, the aforementioned fitness function can be transformed into the following form:

$$\zeta(x) = \begin{cases} A_{\max} - f(x), & \text{if } f(x) < A_{\max}, \\ 0, & \text{else,} \end{cases} \quad (12)$$

$$\zeta(x) = \frac{1}{1 + A_{\max} + f(x)}, \quad (13)$$

where $A_{\max} + f(x) \geq 0$.

2.3.2. Ant Colony Algorithm. The solution to the problem of optimizing the vehicle path is considered the set of feasible paths for ants to find food. The longer the algorithm runs, the higher the pheromone concentration of the shorter path, and the solution to the problem will gradually converge [20]. Therefore, the highest and shortest paths of the pheromone are considered the optimal solution of the optimization problem.

The basic steps of the ant colony algorithm are as follows: first, initialize the parameters and put x ($1, 2, \dots, k, \dots, X$) ants in y ($y_1, 2, \dots, i, \dots, j, \dots, Y$) customers, and the straight-line distance between customer i and customer j is L_{ij} . At time t , the pheromone concentration on the straight path between customers ij is $\vartheta_{ij}(t)$; assuming that the initial time $\Delta\vartheta_{ij}(0) = 0$, the initial iteration number $N = 0$. Then, iterate and visit customers. For each iteration and visit, the number of iterations increases by 1, and the number of ant k also increases by 1. At time t , the state probability of ants moving from customer i to customer j is $P^k_{ij}(t)$, which is calculated as shown in the following formula:

$$P^k_{ij} = \frac{\vartheta_{ij}^\alpha(t) \cdot \mu_{ir}^\beta(t)}{\sum_{r \in \text{allow}_k} \vartheta_{ir}^\alpha(t) \cdot \mu_{ir}^\beta(t)}, \quad r \in \text{allow}_k, \quad (14)$$

$$0, \quad r \notin \text{allow}_k.$$

The ants move to the next customer after selecting the path; calculate the total length of each ant's walking path in this section of the path and find the shortest path. Its function expression is shown in the following formula:

$$\vartheta_{ij}(t+1) = (1 - \rho)\vartheta_{ij}(t) + \Delta\vartheta_{ij}, \quad (15)$$

$$\Delta\vartheta_{ij} = \sum_{k=1}^n \Delta\vartheta^k_{ij}. \quad (16)$$

Ant colony algorithm has many advantages. For example, because it is positive feedback, the convergence speed is fast, and it is performed simultaneously in time [21]. In addition, using the principle of positive feedback, the optimization speed is fast and the optimal solution can be obtained. The disadvantage is that this algorithm is very complicated to solve the problem.

2.3.3. Particle Swarm Algorithm. In the particle swarm algorithm, the birds in the target space are the particles in the solution space, that is, the solution of the algorithm, and the solution to the problem is the food source the birds are looking for. In the

optimization process of the algorithm, the particles always learn from two values: one is the individual historical optimal solution (pbest) and the other is the population historical optimal solution (gbest) [22]. Each particle improves its position and speed based on these two-dimensional values. The size and quality of the fitness value depend on the solution of the actual application problem.

In the particle swarm algorithm, the population size of particles is NP , and each particle can be regarded as a point in the solution space. The position of the first particle is w_i , and the flying speed of the particle is d_i . Therefore, the position and velocity of the particle are updated, as shown in the following formulas:

$$d_{ij}(t+1) = m \cdot d_{ij}(t) + k_1 \cdot r_1 \cdot [pbest_{ij}(t) - w_{ij}(t)] + k_2 \cdot r_2 \cdot [gbest_j(t) - w_{ij}(t)], \quad (17)$$

$$w_{ij}(t+1) = w_{ij}(t) + d_{ij}(t+1), \quad (18)$$

where k_1, k_2 represent the learning factor and are responsible for adjusting the step length of particle learning.

Compared with other optimization algorithms, particle swarm optimization has the same origin as them. All these algorithms need to perform a specific random search in the initial stage of optimization and obtain the next new position through its own iteration. Also, they all use fitness functions to solve individual solutions and global optimal solutions and then use evaluation criteria to obtain the optimal solutions. However, the particle swarm algorithm also has its unique features. In the analysis of the optimization formula, the particle swarm algorithm itself contains the update of speed and position. Although the global search ability is weak, the local search ability and convergence speed are very good.

3. Experiments on Solving Vehicle Routing Optimization Problem Based on Intelligent Algorithm

3.1. Model Solving Algorithm Design Based on Improved GA

3.1.1. Parameter Setting and Coding Design. After setting the parameters, first perform the genetic coding operation on the executable solution of the problem. While generating the initial population, design the fitness function that adapts to the initial population and the termination conditions of the GA according to the input conditions of the mathematical model.

Aiming at the vehicle path optimization problem that this article needs to solve, the natural number sequence is used for encoding. Assuming that x demand points are

delivered by y vehicles, the length of the chromosome at this time is $x + y + 1$, and its code is shown in the following formula:

$$[0, n_{11}, n_{12}, \dots, n_{1r}, 0, n_{21}, n_{22}, \dots, n_{2t}, 0, \dots, 0, n_{y1}, n_{y2}, \dots, n_{yw}, 0] \quad (19)$$

where n_{yw} indicates that the vehicle y provides services to the customer w in the corresponding time period, 0 is the virtual distribution center, and the real distribution center is located before and after 0.

3.1.2. Initial Population and Fitness Function. The fitness function design of this article must first carry out goal transformation and constraint processing. Standardize customer satisfaction and distribution costs and then use linear weighting method to transform; the single-objective function is obtained as

$$Z = \alpha_1 \cdot Z_A + \alpha_2 \cdot Z_B, \quad (20)$$

where α_1 , α_2 , respectively, represent the weights of distribution cost and customer satisfaction ($\alpha_1 + \alpha_2 = 1$), and the weight coefficient can be adjusted according to the actual situation. This paper adopts the penalty function method to deal with the constraint conditions, and the constructed penalty function is

$$F(x, R) = f(x) - R \left[\sum_{i=1}^m \left[\max(0, k_i(x)) + \sum_{j=1}^l r_j(x) \right] \right], \quad (21)$$

where $F(x, R)$ is the penalty function. Based on the above analysis, the fitness function is

$$f = \max_{i=1}^{n+m} Z - R \left[\sum_{i=1}^{n+m} \max(0, x_{ijc}^m \cdot S_i - S, 0) \right], \quad i=1, j=1, c=1, mn+1 \quad (22)$$

where R is an extremely positive number, and other formulas are calculation formulas for parameter constants and variable constraints.

3.1.3. Genetic Operation Design. The selection operation can be performed within an existing executable solution or within the neighborhood of an executable solution. In the GA, for the distribution problem from the distribution center to the customer point, the

selection operation is mainly to find a better solution from the currently executable solution group and use it to replace the original solution group.

The crossover operation uses single-point crossover. The selected two parent individuals are randomly selected as the crossover point.

This article uses the interchange mutation method to perform mutation operations on chromosomes, select chromosomes, and randomly generate two natural numbers; if the natural number is not 0, exchange the genes corresponding to the two natural numbers to generate a new chromosome.

3.2. Experimental Simulation

3.2.1. Experimental Data and Related Parameters. To verify the effectiveness and reliability of the algorithm, this paper conducted a simulation experiment with a distribution center and randomly selected 15 nearby customer points as the distribution nodes. According to the time tolerance interval of each customer point, the time demand sensitivity is evaluated. The demand, expected time window, and tolerance time window of each customer point are shown in Table 1.

The vehicle models are uniform, with a load of 532 kg. The fixed cost for each departure is 120 yuan, and the fuel consumption per kilometer is 0.58 yuan. It is assumed that the vehicle runs at an average speed and the speed is controlled at 45 km/h.

3.2.2. Model Application Testing. Customer satisfaction analysis, according to the customer's expected time window and tolerance time window, analyzes the customer's time, demand sensitivity, satisfaction, and delivery cost.

4. Discussion on Optimization Result of Vehicle Transportation Path

4.1. Algorithm Validity. To ensure the efficiency and probability of GA improvement, in this paper, the number of configurations is set to 300, and the optimal solution results and the average running time of the improved GA, standard GA, and particle swarm algorithm are compared. The convergence diagrams of the three algorithms are as follows.

As shown in Figure 1, the three algorithms have differences in convergence speed. The improved GA in this paper began to converge in the 100th generation, the standard GA began to converge in the 150th generation, and the particle swarm algorithm began to converge in the 160th generation. This shows that the improved GA in this paper is better than the standard GA and particle swarm algorithm in convergence speed. The solution results and running time of the three algorithms are as follows.

As shown in Table 2, the running times of the three algorithms are 37.265 s, 49.338 s, and 51.646 s, the distribution costs are 498.09 yuan, 577.235 yuan, and 598.55 yuan, and the customer satisfaction is 140.45, 117.884, and 116.947, respectively. This shows that

the operating time and optimization results of the GA improvement in this paper are much better than those of the standard GA and particle swarm algorithm.

4.2. Distribution Route Plan

4.2.1. Time Demand Sensitivity. According to the expected time window and tolerable time window of each customer in Table 1, the sensitivity of each customer to the time demand is calculated. The calculation results are as follows.

As shown in Table 3, each customer has his/her own time demand sensitivity and the goods must be delivered within the expected time window; otherwise, compensation will be

Table 1: Customer demand information.

Client	Demand (kg)	Expected time window	Time demand sensitivity
1	120	7:00–8:00	6:30–8:30
2	88	9:30–10:30	9:00–11:00
3	155	9:00–10:00	8:00–11:00
4	45	7:30–8:00	7:30–8:00
5	100	10:00–11:00	9:00–11:30
6	60	9:30–10:30	9:00–11:00
7	135	10:30–11:30	10:00–11:30
8	110	7:00–8:00	6:30–8:30
9	45	11:00–12:00	11:00–12:30
10	36	10:00–11:00	9:00–11:00
11	80	9:00–10:00	9:00–11:00

12	70	8:00– 9:00	7:00– 10:00
13	125	11:30– 12:00	11:30– 12:00
14	120	11:00– 12:00	10:00– 12:30
15	90	8:30– 9:30	7:00– 10:00

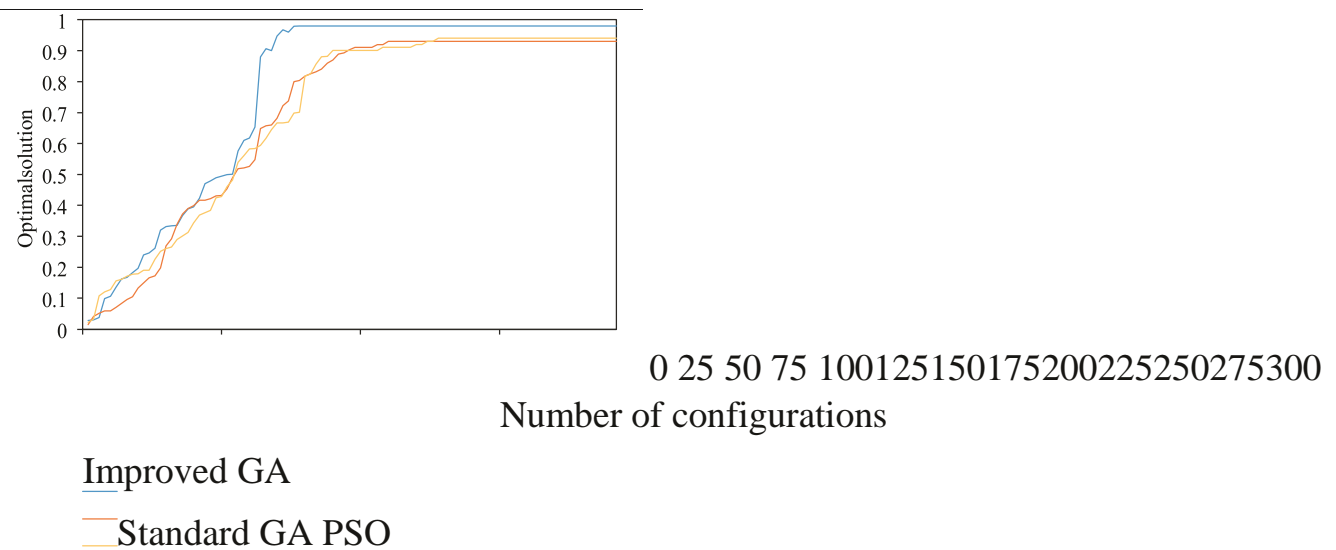


Figure 1: Convergence speed of three algorithms.

required and the distribution cost will increase. To compare the sensitivity of time demand between customers more intuitively, we draw it as a histogram.

As shown in Figure 2, we can intuitively see that customer point 4 and customer point 13 are very time sensitive, and their sensitivity is 1, while customer point 5 and customer point 12 are less sensitive to time. Therefore, it is necessary to plan the distribution plan according to the customer's time sensitivity.

4.2.2. Delivery Route. According to the geographical location and time sensitivity of the customer site, the distribution route is planned. There are three distribution vehicles with the same model, and they start from the distribution center at the same time. The distribution route plan is as follows.

As shown in Figure 3, three vehicles depart from the distribution center at the same time, go to different customer locations, and finally return to the distribution center to indicate the completion of the distribution task.

4.3. Customer Satisfaction and Distribution Costs

4.3.1. Customer Satisfaction. The satisfaction of each customer and each route is investigated, and the distribution plan and the random distribution plan under the improved GA model are used to compare the customer satisfaction of the two. The highest score of each customer satisfaction is 10 points, a total of 150 points. The specific scores are as follows.

As shown in Table 4, there is a big difference in customer satisfaction between the distribution scheme under the GA model and the random distribution scheme. Under the GA model, customer satisfaction can reach a maximum of 10

points and a minimum of 8.9 points. Under the random distribution scheme, the highest customer satisfaction score is 9.5 points and the lowest is only 8 points.

To further analyze the customer satisfaction of different vehicles under different scenarios, we compare the customer satisfaction of the customer points that each vehicle passes. Under the two scenarios, the total satisfaction of vehicle 1, vehicle 2, and vehicle 3 is compared as follows.

As shown in Figure 4, no matter which car, the satisfaction of the delivery plan under the improved GA model is higher than that of the random delivery plan. Especially for vehicle 2, the satisfaction scores of the two are 47.8 and 43 points, respectively, a difference of 4.8 points. After comparing the overall satisfaction of the vehicle, the satisfaction of each customer is compared. The comparison results are as follows.

As shown in Figure 5, although from the overall satisfaction point of view, the distribution scheme under the improved GA model is better, from the perspective of individual customers, not all individual customer satisfaction levels of the random scheme are lower than those of the algorithm scheme. Customer 6's satisfaction with the two solutions is the same (both are 9 points). Customer 8's satisfaction with the random scheme is 9.5 points, which is higher than that of the algorithm scheme by 0.5 points. Customer 15's satisfaction with the random scheme is 9.2 points, which is 1 point higher than that of the algorithm scheme.

4.3.2. Distribution Cost. In the delivery process, the cost composition of each vehicle includes a fixed cost of 120 yuan and a fuel consumption cost of 0.58 yuan per kilometer. If the vehicle arrives earlier than the expected time window, it needs to pay a penalty cost of 0.5 yuan per minute; if the vehicle arrives later than the expected time window, it needs to pay a penalty cost of 0.8 yuan per minute. Through calculation, the distribution cost of the three vehicles under the two scenarios is as follows.

As shown in Table 5, under different distribution schemes, although the fixed cost of each vehicle is the same during distribution, the fuel consumption cost and penalty cost in the actual distribution process are different, so the total distribution cost incurred is also inconsistent. The total distribution cost of the algorithm scheme is 498.09 yuan, and

the total distribution cost of the random scheme is 573.13 yuan. To analyze in detail the distribution cost of three vehicles under different schemes, the distribution cost of each vehicle under the two schemes was compared.

Table 2: Solution results of three algorithms.

Algorithm	Operation hours (s)	Distribution cost	Customer satisfaction
Improved GA	37.265	498.09	140.45
Standard GA	49.338	577.25	117.884
PSO	51.646	598.55	116.947

Table 3: Customer’s time demand sensitivity.

Client	1	2	3	4	5	6	7	8
Sensitivity	0.8	0.8	0.6	1	0.5	0.8	0.8	0.8
Client	9	10	11	12	13	14	15	—
Sensitivity	0.7	0.7	0.6	0.5	1	0.7	0.8	—

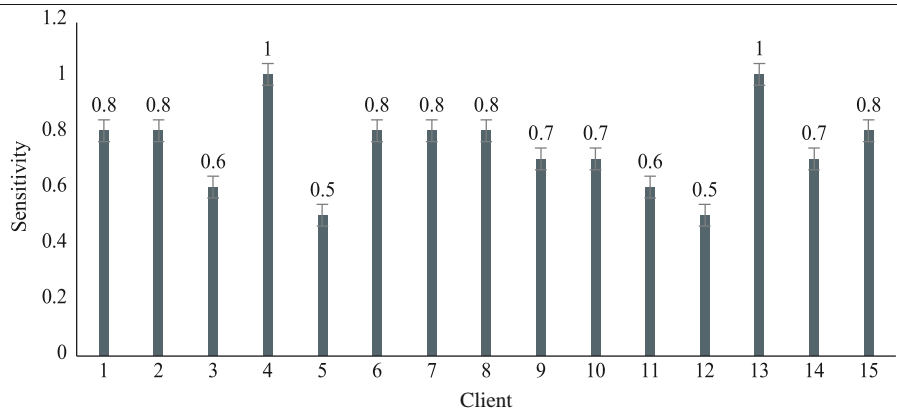


Figure 2: Customer’s time demand sensitivity.

As shown in Figure 6, whether it is fuel consumption cost or penalty cost, the distribution scheme under the improved GA model is almost lower than the random scheme. The biggest difference is the penalty cost of vehicle 3. The difference between

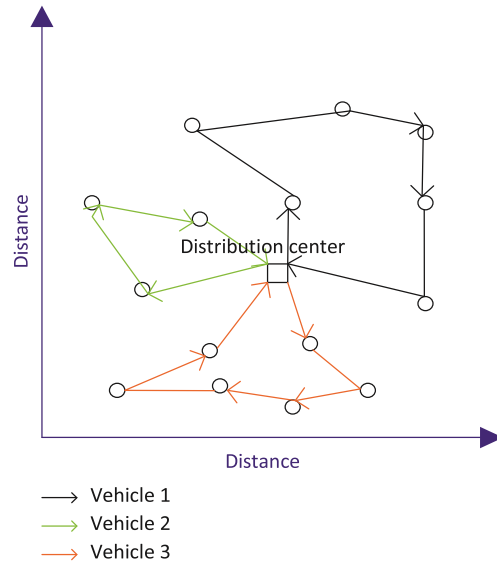


Figure 3: Vehicle distribution path.

the two schemes is 16.07 yuan. Then, we analyze the composition and proportion of the distribution cost of the improved GA, and the results are as follows.

As shown in Figure 7, in the total distribution cost of 498.09 yuan, fixed costs accounted for the largest proportion, totaling 72%, followed by the fuel consumption cost, totaling 20%, and the penalty cost is the least, accounting for only 8%. Among the

three vehicles, vehicle 3 is the one with the most distribution cost, totaling 171.78 yuan, accounting for 35% of the total cost.

Vehicle	Client	Algorithm solution satisfaction	Client	Random scheme satisfaction
1	1	9.5	1	9
	3	10	12	9.1
	6	9	11	8.8
	5	9.45	5	8.75
	9	9.25	9	8.5
Total		47.2	Total	44.15
2	4	9.8	4	9.2
	12	9.7	3	9
	2	10	2	8.5
	7	9.3	10	8.3
	13	9	14	8
Total		47.8	Total	43
3	8	9.45	8	9.5
	15	9.1	15	9.2
	11	9	6	9
	10	9	7	8.5
	14	8.9	13	8
Total		45.45	Total	44.2

Table 4: Comparison of customer satisfaction.

Figure 4: Comparison of customer satisfaction of different vehicles.

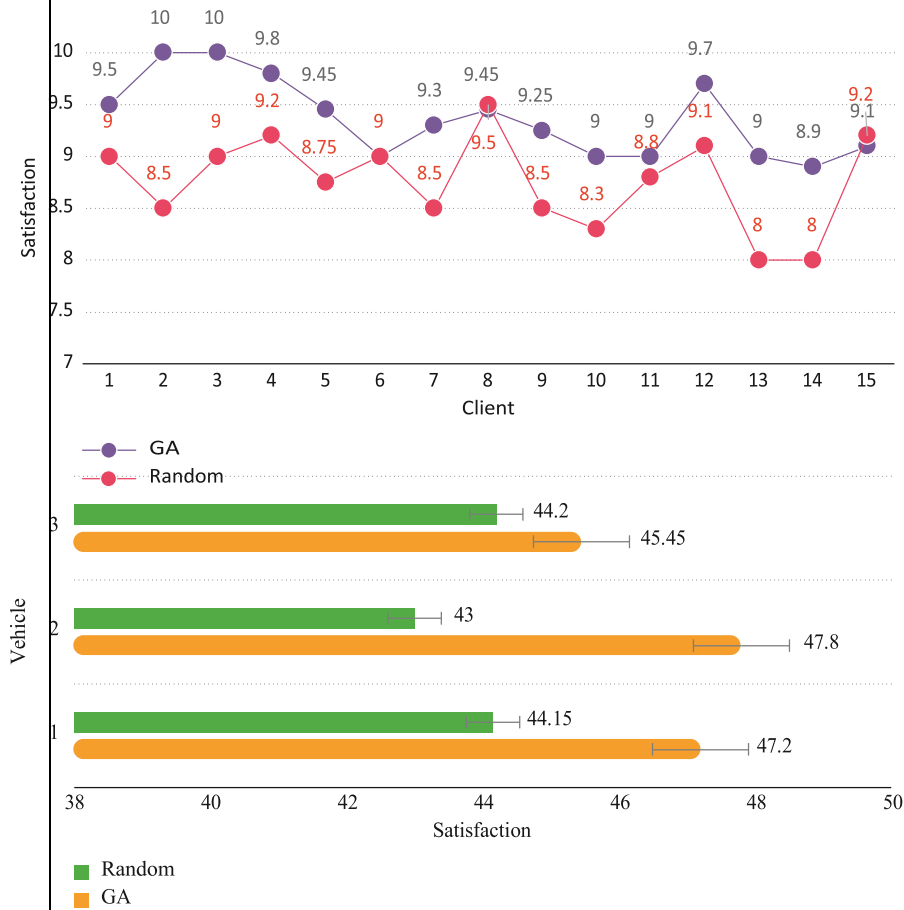


Figure 5: Comparison of satisfaction of different customers.

Table 5: Comparison of vehicle distribution costs.

Vehicle (GA)	Fixed cost	Fuel consumption cost	Penalty cost	Total
1	120	33.65	11.92	165.57
2	120	25.43	15.31	16.74
3	120	38.62	13.16	171.78
Total	360	97.7	40.39	498.09
Vehicle (R)	Fixed cost	Fuel consumption cost	Penalty cost	Total
1	120	48.74	25.24	193.98
2	120	34.01	30.69	184.7
3	120	45.22	29.23	194.45
Total	360	127.97	85.16	573.13

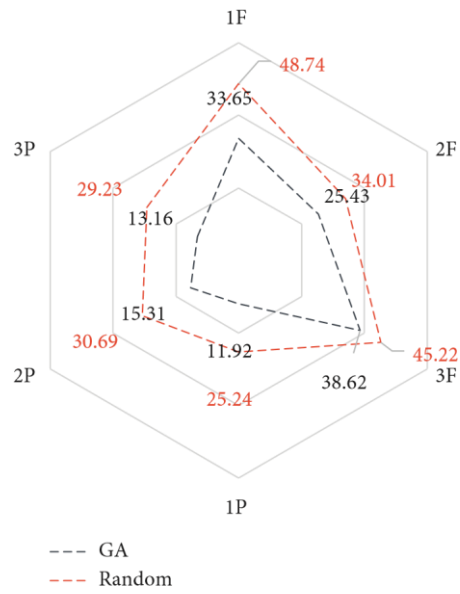


Figure 6: Distribution costs of different vehicles.

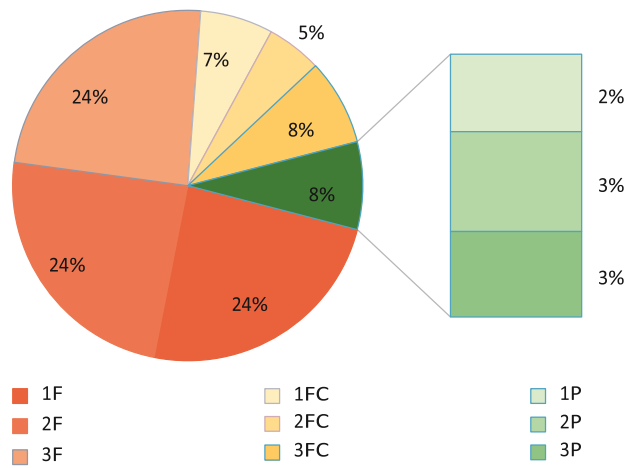


Figure 7: Distribution cost composition proportion under the algorithm scheme.