ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 165
International Conference on Modeling, Simulation, Optimization and Numerical Techniques (SMONT 2019)

Urban Distribution Optimization Based on Order Clustering and Customer Classification

Rui Guo*
School of Economics and Management; Beijing Jiaotong University; Beijing, China
*Corresponding author

***Abstract*—With the improvement of the consumption structure and the increase of consumption levels, the market size of urban distribution is gradually increasing. In response to the problem of large-scale urban distribution vehicle routing problem, a two-stage algorithm of "first partitioning and then scheduling" is proposed. First, the k-means algorithm is used to partition the scale orders according to the order attributes, and then the particle swarm algorithm is used to do the optimization of vehicle routes for the orders in each area; at the same time, the difference in urgency of demands of different customers is considered, and a weight of each customer is introduced to express the priority. Finally, by using D distribution company in Beijing as an example and through programming, the optimal distribution scheduling scheme is obtained. The distribution cost is reduced by 32% and the loading rate is increased by 4%, which verifies the effectiveness and feasibility of the algorithm.**

Keywords—large-scale urban distribution; clustering analysis; customer priority; vehicle routing problem d

I. Introduction

The difference between urban delivery and other types of delivery is that urban delivery has a wide range, scattered orders, and a large customer base. Usually the demand is small but frequent, and the logistics network is complicated. For delivery companies that are targeting supermarkets, the task is even more daunting. Large commercial supermarkets are usually daily-type, with daily distribution requirements and uncertain demand. Large retail companies, such as Wal-Mart, Carrefour, Lotte, etc., ensure that ingredients are fresh, timely, and delivery time requirements are more strict. Therefore, in the urban delivery, whether to meet customer requirements with a high level of delivery service, and to obtain a stable customer base with high customer satisfaction has become an inevitable test for distribution companies.

The existing VRP algorithm can solve the vehicle routing problem with a customer base of less than 100, but for large-scale distribution, due to its time and space constraints, the scale of the problem will be very large. It is difficult for the existing algorithms to get the optimization results quickly and reliably[1, 2]. Baoping Cao(2007)[3] uses the improved base heuristic algorithm to divide the distribution area, and uses the hybrid genetic algorithm to optimize the vehicle route in the area. He R, Xu W, Sun J et al. (2009)[4] uses the traditional k-means method to divide all customers into different regions, and then adjusts the unbalanced regions to be balanced in the second stage by using the boundary adjustment algorithm. Xuedong Gao et al. (2012)[5] proposed a clustering algorithm considering the distribution network structure and distribution volume constraints. Reed, Yiannakou & Evering (2014)[6] used the improved k-means clustering algorithm to divide the customers in waste recovery logistics, and the ant colony algorithm is used to optimize the distribution. Ting Xiang et al (2016)[7] proposed a clustering algorithm which considers vehicle load balance. and used ant colony algorithm for route planning for grouped customers. Daoping Wang et al (2017)[8] adopted a two-stage heuristic algorithm for location selection and vehicle routing optimization of distribution centers. Yang Wang et al(2017)[9] proposed a destruction removal algorithm based on density clustering to obtain a better solution by continuously destroying and rebuilding the current solution.

For large-scale urban distribution order clustering research, it is found that most of the current research is based on the Euclidean distance between customers to partition, while not taking into account the order characteristics of different customers, such as demand and demand time, and so on. With the strengthening of customer personalization, treating all customers as homogeneous customers will greatly reduce customer satisfaction and lower corporate image. Therefore, this paper proposes a two-stage optimization algorithm based on customer classification. After regionalization and integration of customers, customers in different regions are classified according to their own characteristics, such as demand, sales, requirements for distribution services, and so on. Customers with high priority should give priority to distribution. The lower priority customers are placed at the back end of the distribution route, and then the particle swarm optimization algorithm is used to solve the optimal distribution scheme in each region.

II. Mathematical Model

A. Problem Description

The distribution enterprise formulates the daily distribution plan on the basis of the daily business surplus and the demand information (including the type, quantity, time of demand) of the next day, that is, the number of vehicles used and the number of customer points for each vehicle to be delivered. What is the loading capacity of each car? The customer needs and the time window determines the customer delivery priority. How to make the distribution decision efficiently under the consideration of the priority of customer distribution is the main problem to be solved in this paper.

B. Distribution Decision-making Design and Analysis

We need to make the following assumptions necessarily before building model[8]:

1) Only one distribution center. After completing the path tour, the vehicles will come back to the DC.

2) The inventory of DC is sufficient. Each route is delivered by only one vehicle.

3) Each customer node can be served by more than one vehicles, and the efficiency of cargo handling at each customer node is the same.

4) All goods delivered by distribution centerDC are roughly of the same category, and they can be mixed.

5) The demand, geographic location and service time window of the customer nodes are known. The distances between sites and the distances between sites and distribution center are known.

C. The Distribution Decision When All Delivery Tasks are Completed in One Day

Construct the vehicle routing model with the objective of lowest cost. The model description is described as follows:

First introduce variables and symbols.

I : represents the client node set. The customer nodes are represented by the letter i or j.

I_0 : the collection of distribution network nodes where 0 is the distribution center.

K : the collection of vehicles in the distribution center, the vehicles are represented by the letter k said.

d_{ij} : distance from the distribution node i to j.

q_i : the demand of customer node i.

u : the time of loading and unloading of customer nodes.

: the volume of the unit goods.

Q : the rated load of vehicle k.

v : the average traveling speed of the vehicle k.

a_j : the earliest time when the vehicle reach the node j.

b_j : the latest time when the vehicle reach the node j.

t_{jk} : the actual time when the vehicle k arrives at node j.

t_{ijk} : the travel time that the vehicle k travels from the node i to j.

F : fixed start-up cost of vehicle.

C_1 : the transport cost per unit distance of vehicle k.

x_{ijk} =1: vehicle k travels from node i to j; else, x_{ijk} =0.

y_{ki} =1: the task of the node i is done by the vehicle k; else y_{ki} =0

Thus the mathematical model based on the lowest cost is constructed as follows:

Objective function:

$$\min \quad Z=\sum_{j=1}^{n} x_{0jk}\cdot F+\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=1}^{m} x_{ijk}\cdot d_{ij}\cdot C_1 \quad (1)$$

Subject to:

$$\sum_i q_i\cdot y_{ik}\le Q \text{ , } \forall k \quad (2)$$

$$\sum_k y_{ik}=1 \text{ , } \forall i \quad (3)$$

$$\sum_k y_{0k}=\sum_k y_{k0} \quad (4)$$

$$\sum_i x_{ijk}=y_{kj} \text{ , } \forall k \text{ , } \forall j \quad (5)$$

$$\sum_j x_{ijk}=y_{ki} \text{ , } \forall k \text{ , } \forall j \quad (6)$$

$$t_{jk}=t_i+u\cdot q_i+t_{ijk} \quad (7)$$

$$t_{ijk}=d_{ij}/v \quad (8)$$

$$a_j\le t_{jk}\le b_j \quad (9)$$

Objective function: (1) It represents the lowest delivery cost of the scheme. Constraints: (2) It shows that the loading capacity of each vehicle cannot exceed the maximum loading capacity of the vehicle. (3) It represents that the vehicle can only enter the customer node which has been assigned to its delivery task. (4) It indicates that the vehicle can only exit the customer node which has been assigned to its delivery task. (5) It represents the delivery time between the distribution center and customer nodes as well as the customer nodes and customer reaches the client node j. (7-9) are time window constraints.

III. Solution Methodology

A. Analysis of Case

Company D's distribution is mainly intensive urban distribution. Customers send the goods to the warehouse of the company for storage.

The departure time of the vehicles is from 3: 00 to 6: 00 in the morning. The average daily departure vehicles are about 45 vehicles, and the total weight of the goods distributed is 50 to 60 tons. In addition, there are a small number of vehicles delivered in the afternoon, serving Jing Dong and other e-commerce enterprises and undistributed stores in the morning. There is no strict classification of the stores currently required for services, but the large businesses that receive some of the goods strictly (which need to be delivered by 8:00) are classified as category A, requiring priority distribution. The D distribution company now has 15 vehicles of its own, with the same type of vehicle. The rated load is 1500 kg, the unit fixed cost is 500 yuan per vehicle per day, the unit variable cost is 3 yuan per kilometer, and the speed is 80 km/h. A total of 455 orders were extracted from D-distribution companies that needed to deliver one day. The demand for the stores is between 10kg-500kg.

TABLE I. THE DISTANCE BETWEEN THE CUSTOMER NODES AND THE DISTRIBUTION CENTER

d_{ij} **(km)**	**0**	**1**	**2**	**3**	**4**	**5**	**6**
0	0	12487	12569	13122	11939	11701	10625
1	12487	0	336	2932	3461	4850	3945
2	12569	336	0	2599	3176	4575	3738
3	13122	2932	2599	0	1385	2585	2665
4	11939	3461	3176	1380	0	1411	1337
5	11701	4850	4575	2585	1411	0	1512
…	…	…	…	…	…	…	…
380	22766	13698	13378	11102	11658	11316	12702

SPSS software was used to cluster analysis of 380 stores. The cluster area is set to 10 and the number of iterations is 10. The results are shown in the Table II:

TABLE II. CLUSTERING RESULT TABLE

cluster	1	24.000
	2	32.000
	3	34.000
	4	22.000
	5	54.000
	6	38.000
	7	46.000
	8	25.000
	9	44.000
	10	60.000
valid		380.000
invalid		.000

Taking region 4 as an example, the algorithm is used to solve the optimal distribution scheme. The information for each customer point in region 4 is shown in Table III below:

TABLE III. CUSTOMER INFORMATION TABLE OF REGION 4

ID	**Demand /kg**	**Earlier time /min**	**Latest time /min**	**priority**	**distance /km**
91	11.67	60	300	3	12487.17
92	21.11	60	300	1	12569.96
93	38.66	60	300	2	13122.47
94	65.84	60	300	1	11939.26
95	86.89	60	300	2	11701.01
…	…	…	…	…	…
122	40.27	0	120	2	22487.92

B. Algorithm Solving

(1) Determine the service order of the customer node.

We suppose the distribution center has x kinds of business types. And the urgency of each business type is expressed as R_x. 1 is not urgent. 2 represents the general emergency. 3 is very anxious. The weight formula of each customer node is $W = \frac{R \cdot R_x}{D \cdot Demand + T \cdot Time}$. R_x represents the urgency of each business type. $Demand$ represents the demand of each customer node. $Time$ represents the service time window of each client node. R, D, T respectively represents the weight of each parameter. D, T is 1. In order to facilitate comparison of is 10.

Calculate the weight of the customer node and get the distribution point sequence from large to small. For the demand of each customer node, the latest delivery time for each customer node, the distance between each customer point and the distribution center, and the expected quality of service for each customer node. Are the weights of each parameter. Use Java to solve this part of the program as follows:

TABLE IV. THE PROGRAM OF GENERATING VEHICLE WEIGHTS AND SORT

```
for (int i = 0; i < carriers.length; i++) {
for (int j = 0; j < carriers.length - 1; j++) {
float a1 = carsWeight[j];
float a2 = carsWeight[j + 1];
if (a1 >= a2) {
} else {
carsWeight[j + 1] = a1;
carsWeight[j] = a1;
Carrier carrier1 = sortedCars[j];
Carrier carrier2 = sortedCars[j + 1];
sortedCars[j + 1] = carrier1;
sortedCars[j] = carrier2;
}
}
}
carWeight = carsWeight;
return sortedCars;
}
```

(2)The process of particle swarm optimization can be divided into the following steps:

Step1: initializes a group of particles, including random position and velocity.

Step2: evaluates the fitness of each particle.

Step3: compares the fitness value of each particle with the best position it passes through, and if it is better, compares the best position of all particles with pbest.

Step4: as the current best position, and selects the best position in the current population.

Step5: adjusts particle velocity and position according to the above formula.

Step6: does not reach the end condition Step2. The general termination condition is set to a sufficient fitness or to a default maximum iterative algebra.

The solution result is imported into the result.CSV file, and the result of CSV data is analyzed. The results of the example optimization are as follows:

TABLE V. VEHICLE SCHEDULING OPTIMIZATION RESULTS

<table>
<tr><td>The best cost</td><td colspan="2">5658.09</td></tr>
<tr><td>Fixed cost</td><td colspan="2">600</td></tr>
<tr><td>Variable cost</td><td colspan="2">5058.09</td></tr>
<tr><td rowspan="2">The best route</td><td>98-101-95-102-105-108-110-112-93-91</td><td>Loading rate 69.29%</td></tr>
<tr><td>106-99-100-103-104-107-94-109-111-97-96-92</td><td>Loading rate 78.49%</td></tr>
</table>

Similarly, the optimized route information for other regions is shown in Table VI below:

TABLE VI. OPTIMIZED VEHICLE ROUTING RESULTS

region	**Vehicle number**	**The best route**
1	1	6-5-3-1-11-13-22-9-14-10-12-23-16-21-15-17-18-20-19-2-7-4-24-8
2	1	49-53-39-51-25-47-41-46-48-38-29-42-35-45-30-56-36-33-34-50-40-37-52-31-27-28-43-26-32-55-44-54
3	1	82-62-69-80-65-74-58-72-68-75--85-71-67-81-60-59-77-64-63-61-88-86-89-70-90-87-66-79-84-83-78-76-57-73-
5	2	156-147-139-113-115-161-117-120-121-119-138-125-127-140-131-134-144-146-150-153-155-158-159-162-165-124-151-130
		137-143-148-114-116-136-118-132-122-123-135-126-128-129-141-142-145-149-152-154-157-133-160-163-166-164
6	1	180-184-191-178-194-175-188-190-199-197-177-176-192-198-202-185-189-169-182-203-193-195-179-186-201
		212-234-205-210-208-249-238-232-209-215-244-229-218-206-207-217-245-239-221-241-233-220-230
7	1	243-242-228-216-247-227-219-240-231-235-223-214-248-226-225-222-250-211-237-246-236-224-213
8	1	271-269-253-262-251-258-256-259-254-270-263-255-261-264-252-267-265-268-273-266-275-272-260-274-257
9	1	306-311-285-290-277-320-296-310-286-288-308-280-319-307-315-281-283-284-318-297-314-298-293
10	2	338-361-335-368-327-346-331-325-364-321-350-354-370-362-330-366-340-375-374-336-378-357-358-345-323-377
		352-371-359-348-332-324-344-365-341-379-372-353-356-360-376-329-351-373-342-322-328-367-335-363-333-334-343-380-349-339

Because the cost is mainly composed of fixed cost and variable cost, the total cost of distribution before optimization is about 45368.76 yuan, and the cost after optimization is 30658.49 yuan. In the case of the same number of points to be delivered and the same demand for each point, the cost is reduced by about 32.4 knuckles. The loading rate before optimization is about 60, and the average loading rate after optimization is 74.1.

IV. CONCLUSION

This paper studies the vehicle routing problem of large-scale urban distribution based on clustering analysis. The task of distribution is divided into two stages of decision-making. Firstly, the whole distribution area is divided into a geographical location. According to the geographical location of the customers, the customers in the similar areas are divided into the same area. Then on the basis of dividing the completed area, the distribution optimization is carried out in the region. In the process of intraregional distribution optimization, the weight of customer nodes is introduced to describe the degree of urgency of different customers' needs. Then particle swarm optimization algorithm is used to solve the problem. Taking D enterprise of a distribution company in Beijing as a case, the optimal scheme is obtained by solving the case data. In practical application, it increased the loading rate, customer satisfaction and reduced cost. Therefore, this method can provide the reference and basis for the distribution decision.

However, the constraints of actual situation are more complex. In real life, the distance between the customer points is not a straight line, the vehicle will encounter obstacles, single line and other practical problems. The next research can combine the real network to collect the real data of the distance between the customer points. Secondly, in the actual distribution, the attributes of different goods are different, some need to be frozen and refrigerated, some can only be delivered at room temperature, and there is no distinction between them. The follow-up research can start with the property of goods to make the research more practical.

ACKNOWLEDGMENT

This work acknowledges support from Shilin Xu and Zihan Jin. The author is thankful for the valuable comments from the anonymous reviewers.

REFERENCES

[1] Gu W, Foster K, Shang J. Enhancing market service and enterprise operations through a large-scale GIS-based distribution system[J]. Expert Systems with Applications, 2016,55:157-171.

[2] Bin Yu, Penghuan Jin, Zhongzhen Yang. A two-stage heuristic algorithm is used to solve the multi-center vehicle routing problem with time windows [J]. Theory and practice of Systems Engineering, 2012 32 (8): 1793-1800.

[3] Erbao Cao, Mingyong Lai, Kai Nie, et al. Research on vehicle scheduling problem of Large-scale Logistics Distribution [J]. Journal of Hunan University (Auto-science Edition), 2007 / 34 (12): 89-92.

[4] He R, Xu W, Sun J, et al. Balanced K-Means Algorithm for Partitioning Areas in Large-Scale Vehicle Routing Problem: International Symposium on Intelligent Information Technology Application, 2009[C].

[5] Gao Xuedong, Gu Shujuan, Bai Dun, et al. Customer clustering algorithm considering logistics distribution network structure and distribution volume constraint [J]. Theory and practice of Systems Engineering, 2012C 32 (1): 173-181.

[6] Reed M, Yiannakou A, Evering R. An ant colony algorithm for the multi-compartment vehicle routing problem[J]. Applied Soft Computing, 2014,15(2):169-176.

[7] Ting Xiang, Dazhi Pan. A clustering algorithm for solving the demand-divisible vehicle routing problem [J]. Computer applications, 2016N 36 (11): 3141-3145.

[8] Daoping Wang, Zhan Xu, Cen Yang. Research on location-path problem of Logistics Distribution based on two-stage heuristic algorithm [J]. Operations & Management, 2017 / 26 (4): 70-75.

[9] Wang Yang, Guochao He, Yan Wu. A Destruction-removal algorithm for Logistics Distribution problem based on density clustering [J]. Computer applications, 2017, 37 (8): 2387-2394.

[10] Xiaohan Yu, Dong Wang. District partition of urban express delivery based on constrained K-means clustering [J]. Journal of Harbin Commercial University (Natural Science Edition), 2016N 32 (5): 631-634.