An Efficient Density-Based Clustering Algorithm for the Capacitated Vehicle Routing Problem

Jiashan Zhang
Chongqing vocational institute of engineering
Chongqing,402260,China;
zh_jiashan@163.com

***Abstract*—The capacitated vehicle routing problem (CVRP) is one of the most challenging problems in the optimization of distribution. Most approaches can solve case studies involving less than 100 nodes to optimality, but time-consuming. To overcome the limitation, this paper presents a novel two-phase heuristic approach for the capacitated vehicle routing problem. Phase I aims to identifying sets of cost-effective feasible clusters through an improved density-based clustering algorithm. Phase II assigns clusters to vehicles and sequences them on each tour. Max-min ant system is used to order nodes within clusters . The simulation results indicate efficiency of the proposed algorithm.**

Keywords-CVRP; Two-phase heuristic; Density-based clustering algorithm; Max-min ant system

I. INTRODUCTION

The Vehicle Routing Problem (VRP) has been proved to be NP-hard (Laporte 1992). In the past 50 years, hundreds of models and algorithms have been developed to obtain either optimal or heuristic solutions for different versions of VRP, in which the capacitated Vehicle Routing Problem (CVRP) is one of the most famous and widely studied problems. The CVRP involves designing the least cost delivery routes to service a geographically-dispersed customer set, while respecting vehicle-capacity constraints. The majority of current researches focus on the problems within a limited size of 200 customers[1]. Transportation logistics systems are usually large-scale in nature. It is common for real life vehicle routing applications ,such as waste collection, courier service , beverage distribution and milk collection and delivery, to involve the daily service of hundred or even thousand customers. According to the general diagram of the vehicle routing problem, these customers directly are treated as nodes, the street in the city the arc, the scale of the problem will be very large, the difficulty of solving the problem will become greater, the credibility of calculation lower, and the calculation time longer.

The exact algorithms and traditional heuristic algorithms are difficult, even impossible, to solve CVRP. First, the distance in a straight line isn't able to meet problem any longer. Second, calculating the distance matrix is time-consuming. Acctually, besides the distance between customers and the distribution center, the distances among adjacent customers are required, while customers away from each other usually don't belong to the same distribution route and there is little probability of using them. That's to say some (not all) of the distances matrix are used in the process of calculating. So Calculating all the distances between customers are unnecessary .

In the real-life vehicle routing applications, the customers are clustered according to different features, such as road information, customer information, vehicle information, and depot location. Besides simple sweep technology[2], there are several new customer clustering methods. In [3], the customers were firstly devided into districts according to the main road grid system. Then the customer districts were assigned to vehicles using the vehicle flow formulation model. Ouyang [4] proposed algorithms to automatically discretize vehicle routing zones by utilizing a combination of spatial partitioning techniques to systematically obtain optimum zone designs. Ester et al. proposed a density-based clustering algorithm called DBSCAN[5],which is capable of finding arbitrarily shaped clusters. DBSCAN puts nodes with similary density into one cluster, otherwise into defferent clusters. However, in real life distribution, adjacent customers in the same district are seviced by the same vehicle, while customers away from each other are seviced by different vehicles. So, adjacent nodes should be serviced by one vehicle in spite of not reaching the density threshold.

In this literature, CVRP partitions two sub-problems: one is clustering problem ,for which improved DBSCAN is proposed and the other is travel salesman problem (TSP), which is solved by using MMAS.

The rest of this paper is organized as follows. Section 2 introduces the relevant literature. A mathematical programming formulation is developed in Section 3. Section 4 proposes the heuristic algorithm for solving CVRP. Computational results on benchmark instances are reported in Section 5. Finally, conclusions and future work are presented in Section 6.

II. LITERATURE REVIEW

Dantzig and Ramser [6] proposed the CVRP in 1959 at first. Great attention has been devoted on computational experimentation for CVRP since and a variety of algorithms have been developed to solve the CVRP.

Early, constructive heuristics are popular for CVRP. Saving method[7] (Clarke and Wright1964) starts from one dedicated trip for each customer, pairs of trips are merged as long as a saving is obtained. Sweeping method [8](Gillet and Miller 1974)is constructed to generate routes for goods delivery vehicles in which a solution to travelling salesman problem takes place in the second stage of the two stages which exist in Sweep Algorithm. The Mole and Jameson heuristic[9] is another classic in which routes are constructed using successive customer insertions (Mole and Jameson 1976). In general ,they provide solutions at 10-20% above the optimum, in negligible running times.

Tabu search that constituted the most competing algorithms in the 1990s is still present via variants that include sophisticated memory mechanisms. In 1996, Glover [10] presented the advances, applications, and challenges in tabu search and adaptive memory programming. The main idea is to extract a sequence of points (called bones) from a set of solutions and generate a route using adaptive memory. Further , the adaptive large neighborhood search (ALNLS) [11] is presented by Pisinger and Ropke (2007). However, the quality of tabu search depends on the quality of initial solution.

Evolutionary algorithms are proved efficient for the CVRP. [12] presents a grid-based hybrid cellular genetic algorithm for solving the largest existing benchmark instances of CVRP . [13] presents an Parallel Simulated Annealing for large-scale instances. However, the EA is slower than many TS algorithms.

Cluster first-route second methods , proposed by Fisher and Jaikumar [14],is an effective way to deal with CVRP, especially large scale CVRP. It decreases the problem's state space largely . The method first creates customer clusters, each having a total weight not exceeding the vehicle capacity Q and then optimizes the order of visits for each cluster as a TSP subproblem. In the method ,clustering is the key of problem.

III. PROBLEM DESCRIPTION AND FORMULATION

Let $G=(V,E)$ be a complete undirected graph with |V|=n+1 nodes. The node v0∈V represents a depot, where a fleet of m identical vehicles is based, and where the product to be distributed is stored. The other nodes vi∈V\{ v0},for i ∈ {1, ..., n}, represent the customers, characterized by demands for non-negative amounts of product qi. Edges {i,j}∈E represent the possibility of traveling directly from a node (customer or depot) vi∈V to a different node vj∈V for a transportation cost of cij. The CVRP aims to find m or less vehicle routes, i.e. sequences of deliveries to customers, to visit each customer one time exactly while minimizing the total travel distance. The sum of demands should not exceed on any route a value Q assimilated to the vehicle capacity.

The decision variables of the model are:

$$x_{ij}^k = \begin{cases} 1, & \text{if customer j is supplied after vehicle} \\ 0, & otherwise \end{cases}$$

$$y_{jk} = \begin{cases} 1, & if \text{ vehicle k visit} \\ 0, & else \end{cases}$$

A vehicle has a capacity Q, a fixed cost fk and a per unit-distance variable cost gk. The cost of a vehicle of type k traversing the pair (i , j) is denoted by c_{ij}^k ,which is obtained by multiplying the distance dij and the variable cost gk.

The objective function can be written as follows:

$$\min \sum_{k=1}^{m} f_k \sum_{j=1}^{n} x_{oj}^k + \sum_{i=0}^{n} \sum_{j=1}^{n} \sum_{k=1}^{m} c_{ij}^k x_{ij}^k$$

Subjected to

$$\sum_{j=1}^{n} x_{oj}^k = 1, \quad k = 1, 2, K , m \tag{1}$$

$$\sum_{i=1}^{n} x_{ip}^k - \sum_{j=0}^{n} x_{pj}^k = 0, \quad p = 0, 1K\ n; k = 1, 2, K , m \tag{2}$$

$$\sum_{i=1}^{n} q_i y_i^k \le Q, \quad k = 1, 2, K , m \tag{3}$$

$$\sum_{k=1}^{m} y_{ik} = \begin{cases} 1, & i = 1, 2, \ldots, n \\ m, & i = 0 \end{cases} \tag{4}$$

$$\sum_{j=0}^{n} x_{i\,jk} = y_{ik}, \quad i = 1, 2, K , n; \quad k = 1, 2, K , m \tag{5}$$

$$x_{ij}^k \in \{0,1\}, \forall i, j \in V; k = 1, 2, K , m$$

Constraints(1) and (2) state that each vehicle leaves the depot, after arriving at a customer, the vehicle leaves again, and finally returns to the depot. Constraint (3) guarantees that the vehicle capacity will not be exceeded. Constraint (4) and (5) ensure that each client's demand is fulfilled by exactly one vehicle.

IV. IMPROVED DENSITY-BASED CLUSTERING ALGORITHM

Phase I is intended to reduce the computational burden of the subsequent solution phase. By establishing the mathematical model in terms of a few clusters rather than a huge number of customers, the CVRP problem size can be decreased evidently. In this paper, improved density-based clustering algorithm is formulated as follows.

Four input parameters, neighborhood radius ε, the density threshold MinPts and the nearest distance ND, are required and the algorithm also supports the user in determining an appropriate value for the input parameters. They are introduced as follows.

A. E-Neighborhood of a Node

The ε-neighborhood of a node x is defined as

$$N_\varepsilon(x) = \{ y \in D \mid d(x, y) < \varepsilon \}$$

Where ε is neighborhood radius, D is the data set and d(.,.) is a certain distance function.

The Density Threshold : MinPts

Minimum number of points in an ε-neighbourhood of that node.

P belongs to Nε(q) in fig. 1. q is core point only if |Nε(q)| ≥ MinPts.

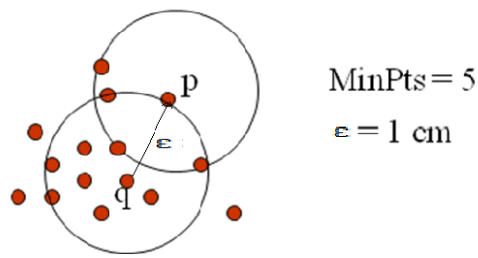

Figure 1. ε-neighborhood and core point

B. The Nearest Distance : ND

Two nodes x,y satisfy the nearest neighbor relationship only if d(x,y)<= ND. ND is constant, Usually ND<ε.

C. Demand Threshold

The total load of a route doesn't exceed the capacity of vehicle. Here, demand threshold W is introduced insuring that the total load of a cluster doesn't exceed W. Generally, w is one-fourth, one-third or half of the capacity q. That is uncertain.

In phase II, the customer clusters were assigned to vehicles using the vehicle flow formulation model. In this paper, saving method is used. So, the vehicle routes are determined as traveling salesman problem(TSP). The detailed routing and scheduling for each tour found is determined by ant colony algorithm (see [15]).

Improved density-based clustering algorithm is described as follows:

Let P be a node∈D, D is the data set. According to neighborhood radius ε and the density threshold MinPts , density-reachable nodes from P or nodes meeting the nearest neighbor relationship merge into one cluster. A cluster is then very intuitively defined as a set of density-connected points that is maximal with respect to density-reachability. If P is core node , nodes which are density-reachable from P or meeting the nearest neighbor relationship , are labelled the same cluster number. Further expansion goes on. If the node P is boundary object and not meeting the nearest neighbor relationship, or total load after merging P exceeds the capacity of vehicle, abandon p and calculate next node. Proceed in order till a cluster produced. Repeat the process till all nodes are labelled. Then, calculation steps into the second phase detailed above.

V. Experiments and Analysis

In this section, we report our computational results. The proposed algorithm has been executed on an Intel Pentium 4 machine with 2GB memory, running windows. Our computational experiment is based on the benchmark instances (1987),see table I.

In the first phase, customers are clustered through the improved strategy. We set $\varepsilon = \bar{D}/3, ND = \bar{D}/4, Minpts = 5, W = q/2$, $\bar{D}$ is the constant ,such as the average distance between customers. In the second phase, MMAS is executed for route scheming. We set $\alpha = 1, \beta = 3, \rho = 0.8$, the number of circulation $N_{c\max} = 200$. The clustering procedure is applied to the instances.

A-n45-k7 has been picked out to detail the computation. The 45 original nodes have been merged into 19 customer clusters(including discrete nodes) , see fig 2. The node size decreased by 56.8%. Cluster C1,C2 and C3 meet the density threshold : MinPts=5, C4, C5, C6, C7, C8, C9 and C10 meet the nearest distance. Others are discrete nodes , not merged into any cluster.

The routes are as follows, see figure 3.

Vehicle 1: 1-32-37-20-30-42-28--1; Vehicle 2: 1-15--6-34-45-25--1;

Vehicle 3: 1--2-38-31-23-11--1; Vehicle 4: 1-13-29-44-12--4--7--1;

Vehicle 5: 1--3--5-22-27-35-36-40--1; Vehicle 6: 1-39-18-26-24-43-16--9-10--1;

Vehicle 7:1-41-21-17--8-19-14-33—1.

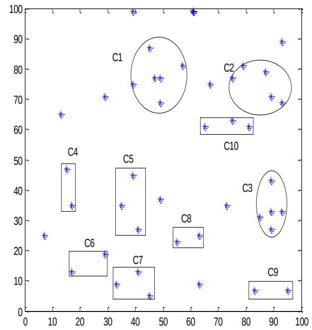

Figure 2. Cluster for problem A-n45-k7

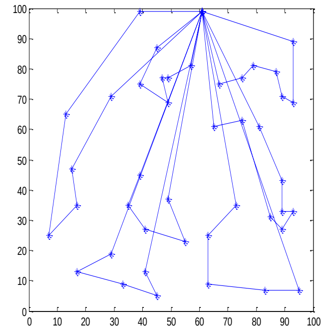

Figure 3. Best solution found for problem A-n45-k7

VI. Conclusion and Future Work

This paper introduces an efficient density-based clustering algorithm for the capacitated Vehicle Routing Problem. The method aims to integrate a heuristic clustering algorithm into an optimization framework.

The method is very successful for clustered examples and solve many of them to optimality. The introduce of preprocessing phase to gather nodes into a few clusters makes the CVRP size decreased sharply. The proposed method can retain optimum in a short time, especially doing well in solving large-scale CVRP. The optimization method is robust, too. Experiments show that density-based clustering algorithm can succeed in solving a variety of benchmark instances.

Real life vehicle routing application is more complicated. For example, the requirement of customers is often uncertain. The extension of the method to these more difficult problems is worth further research.

ACKNOWLEDGMENT

Foundation item: Supported by the National Natural Science Foundation of China (No. 50904032), Chongqing Education Department Science and Technology Project (No. KJ1603201),

Chongqing Association of Higher Education (No.CQGJ15387C) and Chongqing Educational Evaluation Institute (No.PJY2015-50).

REFERENCES

[1] PaoloT, Daniele V. Models, relaxations and exact approaches for the capacitated vehicle routing problem [J]. Discrete Applied Mathematics, 2002, 123: 487-512.

[2] Gillett B, Miller L. A heuristic for the vehicle dispatching problem. Operations Researeh,1974,22:340-349.

[3] Z. W. Qu, L. N. Cai et al, Solution framework for the large scale vehicle de-liver/collection problem, Journal of Tsinghua University (Sci. & Tech.), vol.44, no.5, pp.581-584,2004.

[4] Y. F. Ouyang, Design of vehicle routing zones for large-scale distribution systems,Transportation Research Part B: Methodological, vol.41, no.10, pp.1079-1093, 2014.

[5] Ester M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd Int'1 Conf, on Knowledge Discovery and Data Mining (KDD'96), Portland, Oregon, Aug. 1996, pp. 226-231.

[6] Dantzig, G., Ramser, J.. The truck dispatching problem. Management Science 6 (1), 80-91, 1959.

[7] CIarke G, Wright J.W .Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. Operations Research, 1964,12:568-581.

[8] Gillett B, Miller L.A heuristic for the vehicle dispatching problem. Operations Researeh,1974,22:340-349.

[9] Mole R H, Jameson S R.A . Sequential Route-building Algorithm Employing Generalized Savings Criterion. Operational Research Quarterly,1976,27:503-511.

[10] F. Glover, Tabu search and adaptive memory programming-Advances, applications, and challenges,Interfaces in Computer Science and Operations Research, 1996.

[11] Pisinger D, Ropke S .A general heuristic for vehicle routing problems. Computers & Operations Research 2012,34:2403-2435.

[12] B. Dorronsoro, D. Arias, A Grid-based hybrid cellular genetic algorithm for very large instances of the VRP. Parallel and Grid Computing for Optimization, PGCO 2007.

[13] Czech Z. J, Czarnas P. Parallel Simulated Annealing for the Vehicle Routing Problem with Time Windows. Proceedings of the 10th Euromicro Workshop on Parallel, Distributed and Network-based Processing, 2015: 376-379.

[14] Fisher, M., Jaikumar, R.. A generalized assignment heuristic for vehicle routing. Networks 11 (2), 109-124, 1981.

[15] T. Stützle, H.H. Hoos, The MAX-MIN ant system and local search for the traveling salesman problem, Proceedings of the IEEE International Conference on Evolutionary Computation (ICEC97), IEEE Press, Piscataway, USA, 1997, pp. 309-314.

TABLE I. SUMMARY OF COMPUTATION FOR BENCHMARK INSTANCES

Problems instance	**Node_ number**	**Vehicle_ number**	**Computed cost**	**Cluster_ number**	**Node _size reduction**	**Vehicle_ number**	**Best Known cost**	**Deviation percentage(%)**
A-n37-k5	36	5	709	21	41.7%	6	669	6.0%
A-n37-k6	36	6	980	22	38.9%	6	949	3.3%
A-n45-k7	44	7	1192	19	56.8%	7	1167	2.1%
A-n54-k7	53	7	1227	33	37.7%	7	1167	5.1%
A-n63-k9	62	9	1714	41	33.9%	9	1616	6.1%
A-n63-k10	62	10	1410	39	37.1%	11	1314	7.3%
E-n30-k3	29	3	527	17	41.4%	3	508	3.7%
E-n33-k4	32	4	858	19	40.6%	4	837	2.5%
E-n51-k5	50	5	538	28	44.0%	5	524	2.7%
E-n76-k7	75	7	713	43	42.7%	7	687	3.8%