Acta Technica *62* No. 4B/2017, 1–8 © 2017 Institute of Thermomechanics CAS, v.v.i.

Finding optimized routes in vehicle routing problem with backhaul by combining genetic algorithm and improved K-means

Saeed Mohammadian-Semnani[1,2], Mahdieh Karkehabadi[1,3]

Abstract. Vehicle routing problem in road network is one of the most widely aspects of load displacement. This problem can play a very important role in efficiency of country cargo fleet transportation. This study considers the backhaul problem in vehicle routing which is one of the major subcategories of vehicle routing problem. Vehicle routing problem with backhaul is a development section in vehicle routing problems which consists of backhaul and linehaul customers. For the first time, in one of the current researches, a novel hybrid K-means and metaheuristic, based on the bio-inspired genetic optimizer is presented to solve VRPB, the Vehicle routing problems with backhaul. In the first phase of the research, customers are divided into a number of feasible groups; in the second phase, the best route is determined to visit all customers. Finally, we will use genetic algorithm to find the best possible way for constructing route within each cluster and between two clusters in two separate sets of customers. The result of algorithm performance was used to deal with a number of benchmark problems; they indicated that this method has acceptable results and is quicker in comparison with the other algorithms.

Key words. Vehicle routing, metaheuristic, improved K-means, genetic algorithm.

1. Introduction

Optimization packages have been widely used in the past decades, based on Operations Research and Mathematical Programming techniques, to effectively manage the provision of goods and services in distribution systems. The significant impact of this savings on the global systems of economy is evident. In fact, all the production procedures and distribution systems are included in transportation process which

[1]Faculty of Mathematics, Statistics and Computer Science Semnan University, Semnan, Iran
[2]E-mail: `S.Mohammadian@semnan.ac.ir`
[3]E-mail: `M.Karkeabadi@hotmail.com`

represents a component related to the final cost.

One of the most important distribution problems, are the ones which concern the goods distribution between depots and final users (customers). These problems are generally known as Vehicle Routing Problems (hence forth VRPs) or Vehicle Scheduling Problems.

These problems have a central location and divide to several branches, each of which has a various application in distribution systems. We consider one of the main branches of Vehicle Routing problem where, besides constraints of VRP, other constraints must also be active to help us find the solution. This branch of VRP deals with backhauls, so it is important for return route and loss of routes without load. In VRP with Backhauls (hence forth VRPB), customers are divided into two groups: linehaul customers and backhaul customers.

A linehaul customer needs a certain amount of goods for receiving, while backhaul customer should deliver a certain amount of goods. The goal in this problem is to design a set of minimum routes in a way that the both groups' total load of costumers does not exceed from the capacity of vehicles in each route. As a branch of VRP, VRPB is also considered as an NP-hard problem, because by considering $B = \emptyset$ this problem converts to VRP [1–3].

The branch and bound algorithm are of the exact methods which are unable to solve large size problems, while simple heuristic approaches provide limited quality of the solution [4]. The metaheuristic algorithm is meant to fill in this gap for the simple heuristics through relaxing the restriction of computation time. Although a metaheuristic may be more time-consuming for CPU, it would yield a better quality solution. Metaheuristic methods improve heuristic methods by minimizing computation time. A metaheuristic method may prolong the CPU time, but create a solution with high quality. This paper is organized as follows: In Section 2 we bring a review of VRPB. After that in third section there is a review of Clustering. Section 4 presents genetic algorithm and operators used in problem solving. In Section 5 we proposed a modeling of VRPB with Clustering and metahuristic. In Section 6, a computational experiment with some of benchmark problems available in the literature and the comparison of the other metahuristic's performance with proposed method is presented, and finally, Section 7 consists of conclusions and informs about future works.

2. VRP with Backhauls

The VRP with Backhauls (VRPB) is the extension of the CVRP where the customer set V_0 is divided into two subsets. The first one, L, holds n linehaul customers, each calling for a given quantity of product to be delivered. The second subset, B holds$m$ backhaul customers, in which inbound product requires a given quantity to be assigned [5]. Customers are indexed for $L = 1, \dots, n$ and $B = n+1, \dots, n+m$.

In the VRPB, linehaul and backhaul customers observe a precedence constraint between them: by serving both types of customer, the route first deals with all the linehaul customers before any backhaul customer then moves on to them as proceeding. A nonnegative demand, d_i, is related to each customer i, and the depot is associated with a fictitious demand $d_0 = 0$. This demand is to be either delivered or collected based on its type,

Utilizing the graph theory, the definition of the problem can be presented as following. Consider a complete undirected graph $G = (N, A)$, where $N = \{0\} \cup L \cup B$ is a set of $l+b+1$ vertices and subsets L and B represent the two customers, linehaul and backhaul. As a set of arcs, $A = \{(i, j)\ ,\ i, j \epsilon N\}$ is associated with arc (i, j), where a nonnegative cost $C\,(i, j)$ exists. In case of cost matrix being asymmetric, the problem will be also asymmetric VRPB.

VRPB includes obtaining a collection of exactly K simple circuits with minimum cost, under these conditions:

(I) The depot vertex is met by each circuit.

(II) Exactly one circuit visits each customer vertex.

(III) The vehicle capacity C must be observed for the over-all demands of the linehaul and backhaul customers being visited by a circuit.

(IV) In each circuit, in case of presence of any backhaul customers, they must go after the linehaul customers.

Let's define $L_0 = L \cup \{0\}\ ,\ \ B_0 = B \cup \{0\}$. Let $\overline{G} = (\overline{V{,}A})$ be a directed graph obtained from G by defining $\overline{V} = V$ and $\overline{A} = A_1 \cup A_2 \cup A_3$, where three disjoint subsets can be driven from partitioning the arc set A. These three subsets are shown as follows:

$A_1 = \{(i, j) \in A : i \in L_0,\ j \in L\}$, $A_2 = \{(i, j) \in A : i \in B,\ j \in B_0\}$, and $A_3 = \{(i, j) \in A : i \in L,\ j \in B_0\}$.

Set $\overline{A}$ contains all the arcs which belong to a feasible solution. Let $\mathcal{F} = \mathcal{L} \cup \mathcal{B}$, where $\mathcal{L}$ and $\mathcal{B}$ are considered as the family of all subsets of vertices in L and B respectively. Moreover, for each $S \in \mathcal{F}$, let $r(S)$ be the minimum number of vehicles needed for serving all the customers in S. For each $i \in V$, let us define $\Delta_i^+ = \{j : (i, j) \in \overline{A}\}$, $\Delta_i^- = \{j : (j, i) \in \overline{A}\}$. X_{ij} is the variable of model which takes the values of 0 and 1.

3. Clustering (K-Means method)

This method, despite its simplicity, is a basic approach for many other clustering methods (such as Fuzzy Clustering). The algorithm of this method has been expressed in different ways but all of them have repetitive sequences which try to estimate the following issues for a fixed number of clusters:

1. Obtaining points as (these points are in fact the average of points in each cluster).

2. Attributing each data sample to a cluster with minimum distance to the cluster center.

By repeating the same procedure, we can calculate new centers for data in each iteration obtaining the average of data and attribute data to new. This process continues up to discontinuation of changes in the data. The objective function is presented as $\sum\sum\left\|X_i^j - C_j\right\|^2$ and indicates the criterion of distance between points and C_j (the center of the cluster j).

The following algorithm is a basic algorithm for this method:

Step 1: Choose k points as cluster centers points.

Step 2: Each data sample attributes to the cluster whose center has the minimum distance from that data.

Step 3: After belonging all data to one of the clusters, a new point is calculated as the center point (the average of points belonging to each cluster).

Step 4: Redo steps 2 and 3 until no changes would occur in center clusters.

4. Solving vehicle routing problem with backhauls using clustering and GA

Since the number of variables 0 and 1 in Routing problems is large, they belong to NP-hard problems which make us use heuristic and meta-heuristic algorithms to solve them; however, we might reach a solution away from the main answer because of using probability rules in these algorithms. Although in our proposed method we used meta-heuristic algorithm for prioritization of each customer in each cluster, it is better to solve problem with large number of data with proposed method. But, due to the fact that data contained in a cluster are low and very close to each other, the solution will not be out of mind.

Now, according to the definition of problem we know that the set of customers are partitioned into two sets L and B. To solve the problem with proposed method we act as bellow: First, each of partitions are considered as an independent set, so we consider the set consisting of linehaul (L) in first step and cluster the customers in it, based on euclidean distance from each other and vehicles capacity. If we only consider the distance between customers, it may lead to form extra clusters or vehicles may not be allowed to service a cluster.

In next phase, when all customers of two groups were clustered, we consider vehicles from depot for servicing based on the number of clusters in set L so that the condition of capacity constraints of vehicle is active for becoming dependent to clusters. This condition expresses the fact that the demands for all customers in a cluster should not be more than the capacity of vehicle. After that, for each vehicle related to each cluster in set L, we select the best cluster from set B for depot according to the conditions such as vehicle capacity and supplying of goods of that cluster. To construct route between two groups we can use any of the meta-heuristic methods to create the best possible way (the minimum distance). Moreover, to have the best servicing for customers in each cluster, we can use optimization meta-heuristic methods for prioritization (to find an optimal solution in terms of traveling less distance and saving time). This causes the vehicles to service the customers according to the priorities and therefore travel less distance in less time for service delivery.

5. Computational results

In this section, we compared the solutions obtained from the proposed method with the solutions of several methods that have been used previously to solve this problem. The numerical experiment was proposed by Goetschalckx and Jacobs-Blecha (1989) [2] that we used some of these problems for proposed algorithm. The examples vary in size from 25 to 150 customers with backhauls ranging between 20 and 50 %. In problem instances for the x coordinate, customer coordinates are homogenously distributed in interval [0, 24 000], and in interval [0, 32 000] for the y coordinate. The depot is located in the center at ($x = 12\,000, y = 6\,000$). The results of solving the problem were acceptable. The algorithm is applicable for any number of customers. The solutions found of implementation of the algorithm on some of the benchmarks are shown in Table 1. Algorithm owns appropriate efficiency in some of the small problems. In these problems, there is no difference between the solution and the best solution found thus far. The solutions were obtained with the same quality in the better time. In some of big problems, the solution is a little different from the best solution found so far, but the algorithm implementation time is less. In all of the problems presented, the time of algorithm implementation is faster rather than the other metaheuristic algorithms.

Table 1. Total tour length obtained

Total customer	Linehual customer	Backhaul customer	Vehicle capacity	Vehicle number	Our length
30	20	10	1600	7	239080.16
40	20	20	4150	5	199345.96
38	30	8	1700	11	316708.86
57	45	12	5300	5	229507.48
68	45	23	6100	5	250220.72
90	45	45	4000	7	309943.84
94	75	19	4400	10	335007.69
150	75	75	4400	10	439700.81
125	100	25	6200	9	378159.37

In the following, the solution for a problem with 125 customers is shown. Customers are drawn in original coordinates. Figure 1 shows the distribution of customers in this problem.

6. Conclusion

In this paper we developed a combination of improved K-means classification and Genetic Algorithm. It is then used for the first time to solve the problem of Vehicle

Routing problem with backhauls.

Table 2. Comparison of proposed algorithm with the best solution known for VRPB

The best solution known	Our length (m)
239080.16	239080.16
199345.96	199345.96
316708.86	316708.86
229507.48	229507.48
250220.77	250220.77
309943.84	309943.84
418828.93	439700.81
377328.75	378159.37

Table 3. Comparison time of proposed algorithm with the best solution known

The best solution known	Our time (s)
2.13	0.071
4.88	1.241
6.25	2.366
14.25	9.811
27.13	15.743
37.50	21.976
66.63	40.718
148.38	101.689
183.75	128.459

Improved K-means is used for classification (clustering) of customers in two different set. This improved method maintains the limitation of vehicle capacity in each cluster in comparison with K-means method. GA is used for prioritization of services in each cluster as well as making route between two clusters in two separate sets of customers (linehual and backhual). Metaheuristic algorithms shows better performance (result) in small spaces so with clustering the customers in each group, the number customers for search will reduce so better results will be achieved. The advantage of this method is that it is faster than the other heuristic algorithms and the results are reasonably acceptable. This algorithm can be used for other sub-branches VRPs. We can use metaheuristic in the initial clustering and choose the best center for each and then it implements on this problem and the solutions of both methods compared.

References

[1] J. Brandão: *A new tabu search algorithm for the vehicle routing problem with backhauls.* European Journal of Operational Research *173* (2006), No. 2, 540–555.

[2] M. Goetschalckx, C. Jacobs-Blecha: *The vehicle routing problem with backhauls.*

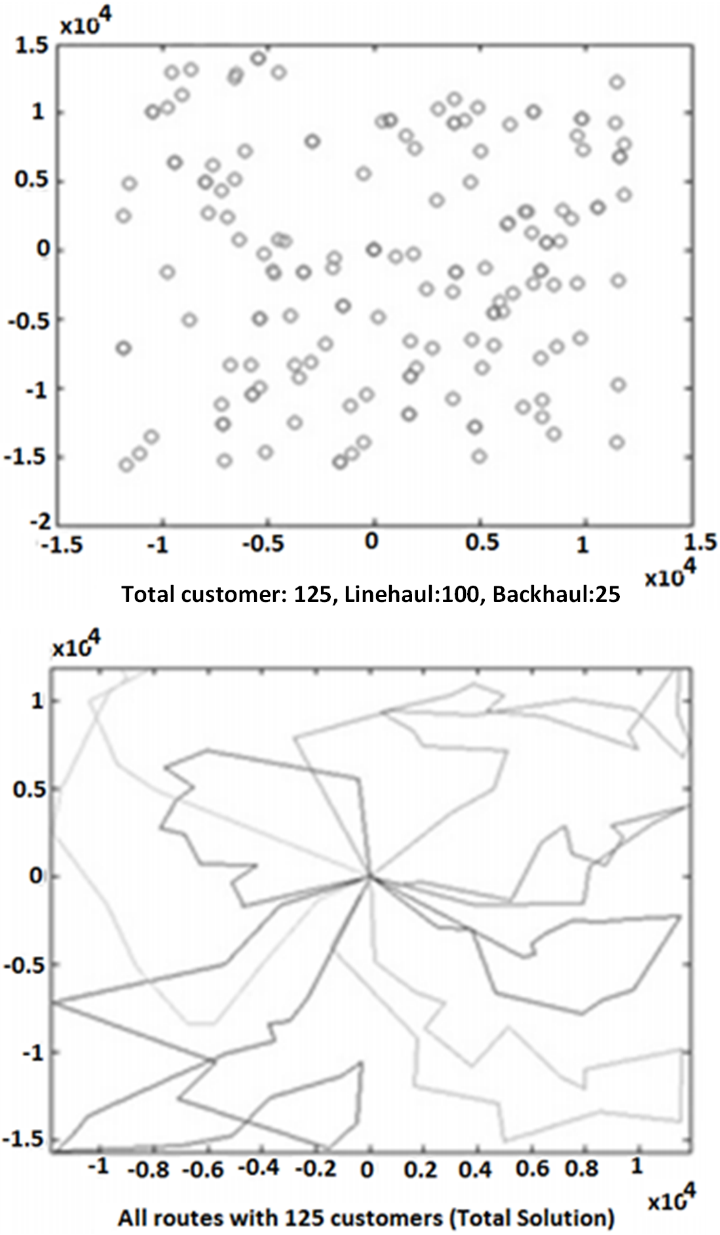

Fig. 1. Distribution of customers in the problem

European Journal of Operational Research *42* (1989), No. 1, 39–51.

[3] A. K. Jain, M. N. Murthy, P. J. Flynn: *Data clustering: A review.* Journal ACM Computing Surveys (CSUR) *31* (1999), No. 3, 264–323.

[4] I. H. Osman, N. A. Wassan: *A reactive tabu search meta-heuristic for the vehicle*

routing problem with back-hauls. Journal of Scheduling *5* (2002), No. TOC 4, 263–285.

[5] P. L. Abad, Y. Gajpal: *Multi-ant colony system (MACS) for a vehicle routing problem with backhauls*. European Journal of Operational Research *196* (2009), No. 1, 102–117.

[6] M. Reimann, K. Doerner, R. F. Hartl: *D-Ants: Savings based ants divide and conquer the vehicle routing problem*. Computers & Operations Research *31* (2004), No. 4, 563–591.

[7] A. C. Wade, S. Salhi: *An ant system algorithm for the mixed vehicle routing problem with backhauls*. Metaheuristics: Computer Decision-Making, Applied Optimization, Book Series (APOP), Kluwer Academic Publishers *86* (2004), 699–719.

Received October 31, 2017