Solving the Heterogeneous Capacitated Vehicle Routing Problem using K-Means Clustering and Valid Inequalities

Noha Mostafa
Department of Industrial Engineering and Systems Management
Egypt-Japan University of Science and Technology
New Borg El Arab, Alexandria 21934, Egypt
noha.ali@ejust.edu.eg

Amr Eltawil
Department of Industrial Engineering and Systems Management
Egypt-Japan University of Science and Technology
New Borg El Arab, Alexandria 21934, Egypt
eltawil@ejust.edu.eg

Abstract

The Capacitated Vehicle Routing Problem (CVRP) is one of the most popular routing problems. In the CVRP, a fleet of capacitated vehicles located at a central depot are used to deliver products to a set of geographically dispersed customers with known demand. In this work, a new two-step approach to solve the CVRP with a heterogeneous fleet is proposed. In the first step; a balanced K-means clustering algorithm is used to aggregate the customers into balanced clusters. In the second step, clusters are assigned to vehicles and a MIP model is solved with adding valid inequalities that are used to reduce the volume of the mathematical model. Then the problem is disaggregated to find the detailed tours for each vehicle. The proposed approach was used to solve benchmark problems from literature, and it proved its efficiency in terms of quality, vehicle utilization, and computational time. The contribution of this work is combining the clustering algorithm with cutting techniques in order to find near-optimal solutions within reasonable computational times.

Keywords
K-means clustering, valid inequalities, vehicle routing problem, supply chain management

1. Introduction

The research on the Vehicle Routing Problem (VRP) started about sixty years ago when Dantzig and Ramser published their famous paper 'The truck dispatching problem' (Dantzig and Ramser, 1959). Since then many researchers targeted this problem considering different variations and solution approaches. The main objective of the VRP is to find the best routes from a single or multiple depots to a set of customers by using a single or multiple vehicles, subject to some constraints with the objective of minimizing the total travelled distance or total travel cost. Many variants exist for the VRP, the most famous problems are the Capacitated VRP (CVRP) (Toth and Vigo, 2014), VRP with time windows (Baldacci et al., 2012), VRP with pickup and delivery (Wassan and Nagy, 2014), the periodic VRP (Campbell and Wilson, 2014), VRP with multiple depots (Montoya-Torres et al., 2015), the dynamic VRP (Pillac et al., 2013), and the green VRP or the Pollution routing problem (Bektas and Laporte, 2011). The VRP is a combinatorial optimization NP-hard problem. So, exact algorithms are possible only for small-sized instances (Garey and Johnson, 1979). Hence, most of the works resort to heuristics and metaheuristics. For recent reviews on the VRP and its solution methods please refer to Braekers et al. (2016) and Abdelhalim and Eltawil (2012).

In this paper, a new solution approach for the heterogeneous CVRP problem is proposed. In order to deal with the complexity of the problem, it is necessary to take some steps to cut down the computational runtime; the first step is to apply partitioning-based clustering for the customers, and then adding valid inequalities to the model to make an

approximation of the convex hull of the polyhedron of the problem, and hence reduce the computational time for running the model.
Customers' clustering is used to group the customers into logical segments according to some criteria so that it is more economic and/or less complicated to optimize the logistics function and hence the supply chain performance. Customers clustering and geographic zones partitioning have become an important issue in supply chains especially in the case of Fast-Moving Consumer Goods (FMCG) distribution systems. Many criteria or performance indicators can be considered before performing the clustering, it depends on the type of the supply chain, the type of product(s), facilities locations, fleet size, etc. The most common criteria are the total route transportation costs, total distance, route duration, workload balancing, and time or delivery constraints. Other criteria can take into consideration demand patterns and environmental issues. Clustering decisions, in general, can be strategic, tactical or operational. An example of a strategic clustering decision is deciding the political zones for elections, planning of public schools and hospitals, supporting decisions for supply chain location problems etc. Tactical clustering decisions can include emergency services shelters, garbage collection segmentation. Finally, an obvious example on operational clustering decisions is delivery and or pick up operations. Hence, customer clustering can be a tactical or an operational decision according to the time span of decision making and the frequency of change that is related to the dynamicity of the supply chain itself. Valid inequalities are inequalities for a set that are satisfied by each point in the set (Cornuejols, 2008). They are useful when the polyhedron of the model is complex, so using cuts can reduce its hypervolume and consequently the required computational effort. There are many techniques that can be used to generate cuts, the most common are: Lift-and-Project, Gomory Mixed Integer cuts (GMI), K-cuts, reduce-and-split cuts, Chvatal-Gomory cuts (CG cuts), and Disjunctive cuts (Cornuejols, 2008; Balas et al., 2013).
The rest of this paper is organized as follows: In section 2, a literature review is given on using clustering and valid inequalities for supply chain problems. Section 3 illustrates the methodology used in this work. Section 4 gives the computational results and Section 5 draws the conclusions and future work.

2. Literature Review

2.1 Clustering in supply chains

Clustering or cluster analysis was mainly developed as an analysis tool for statistical data; it was used in many applications such as data mining, pattern recognition, machine learning, and bioinformatics. Frohlich and Westbrook (2001) showed that using cluster analysis can have a positive impact on supply chain management. Researchers started to use cluster analysis for different supply chain areas. Early works were directed towards item partitioning in inventory (Axsater, 1981; Mitchell, 1983), production (Ernst and Cohen, 1990), location (Hwang, 2002) and E-business (Cagliano et al., 2003). However, a common application of clustering in supply chain was to perform customers clustering for vehicle routing. The work by Keeney (1972) recommended using clustering to assign customer areas to different facilities, and after that many researchers targeted this problem by using exact methods (multi-objective optimization, mixed-integer programming, and continuous modelling) or non-exact methods (division or agglomeration heuristics).
In the recent ten years, researchers paid more attention to approximation (non-exact) methods due to the increased complexity of case studies and large data sets. Miranda and Garrido (2004) used Lagrangian relaxation heuristic to solve a hub-and-spokes model that assigns fleet clusters to a set of distribution centers. Dondo and Cerda (2007) proposed a clustering algorithm and used it as a pre-processing stage in solving a vehicle routing problem with time windows, a heterogeneous fleet of vehicles, and multiple depots, by using this clustering algorithm they could solve a problem with up to 100 customers instead of 25 without using clustering. Barreto et al. (2007) proposed a cluster analysis based sequential heuristic in a capacitated location routing problem, they used four grouping techniques and six proximity measures. Their results showed a big improvement after applying the clustering with an average gap of 4.81%. Mitra (2008) proposed a parallel clustering and route construction heuristic for a vehicle routing problem with split deliveries and pickups in which the number of clusters is known a priori.
Yücenur and Demirel (2011) proposed a geometric shape based genetic clustering algorithm to solve a multi-depot vehicle routing problem, they compared this algorithm with a classical nearest neighborhood and shown the superiority of the proposed algorithm. Erdogan and Miller-Hooks (2012) used a three-step density-based clustering algorithm to solve a vehicle routing problem with environmental considerations. In this algorithm, the neighborhood of each vertex in a cluster with a given radius must contain at least a minimum number of vertices. After the clustering is done, a route is constructed for each vehicle and then the best set of routes is selected according to the least total distance.

K-means clustering has been a typical approach used in many recent papers. He et al. (2009) used a two-step K-means algorithm. In the first stage all the customers are partitioned in several clusters and then in the second stage a border adjustment algorithm is used to make the number of customers balanced between the clusters. Sahraeian and Kaveh (2010) proposed a hybrid algorithm that uses K-means clustering and fixed-neighborhood search to solve a discrete capacitated P-median location problem. Lin et al. (2013) used a data mining engine to support operational decisions for the routing in a city logistics scope, K-means algorithm was used as an independent module to cluster the customers. Nananukul (2013) developed an enhanced clustering model that include a K-means algorithm to solve a clustering model for customers that considers demand patterns and holding costs.
Most recently, Praveen et al. (2016) used an improved k-means clustering algorithm that can enhance the classical saving matrix method to minimize the distance and the number of vehicles in a CVRP. Cinar et al. (2016) proposed a two-phase constructive heuristic that includes a K-mean clustering algorithm to solve a cumulative vehicle routing problem with limited duration. Ferrandez et al. (2016) used K-means clustering to find launching locations for a truck-drone network, followed by a genetic algorithm to solve the routing problem.

2.2 Valid inequalities in supply chains

Valid inequalities have been used in many supply chain problems to find better lower bounds. For example they were used for the inventory-location-routing problem (Guerrero et al., 2013; 2015), the inventory routing problem (Bertazzi et al., 2013; Coelho and Laporte, 2013; Agra et al., 2013), the location-routing problem (Doulabi and Seifi, 2013; Contardo et al., 2014), the production-distribution problem (Rizk et al., 2008; Melo and Wolsey, 2012), and the vehicle routing problem (Dror et al., 1994; Belenguer et al., 2000; Jin et al., 2007; Perboli and Tadei, 2010).

3. Problem description and methodology

The CVRP with a heterogeneous fleet is modelled as a Mixed Integer Programming model (MIP), and then an enhanced K-means clustering method is used to cluster the customers and valid inequalities are added to the MIP model.

3.1 The Proposed Mathematical model

The model is represented by an undirected graph with set N of nodes and set E of edges; $N = \{0, 1, \ldots, n\}$ where node 0 represents the depot, and $E = \{(i, j): i, j \in N, i \neq j\}$ where i and j represent a set of geographically dispersed customers. There is a set K of products, a set T of planning periods and a set V of heterogeneous vehicles that are used to deliver the products to customers and then return to the depot. The demand of each customer is known beforehand; d_{ikt}. Each vehicle has a capacity q_v that cannot be exceeded. The distance between nodes i and j by using vehicle v is r_{ijv}. Shortages are not allowed, so the demand of each customer should be fulfilled. The decision variables are:

$$W_{it} = \begin{cases} 1 & \text{if customer } i \text{ is visited in period } t \\ o & \text{otherwise,} \end{cases}$$

X_{iktv} = Amount of product k delivered to customer i by vehicle v in period t

Y_{ijktv} = Amount of product k delivered between nodes i and j by vehicle v in period t

$$Z_{ijtv} = \begin{cases} 1 & \text{if vehicle } v \text{ is used from node } i \text{ to node } j \\ & \text{in period } t \\ o & \text{otherwise.} \end{cases}$$

The model was formulated as a mixed-integer programming model that focuses on partitioning the customers and deciding the amounts to be transported over the planning horizon. The objective function is to minimize the total distance for all clusters:

$$Minimize \sum_{t \in T} \sum_{v \in V} \sum_{i \in N} \sum_{j \in N\{i\}} r_{ijv} Z_{ijt} \tag{1}$$

$Subject\ to$

$$\sum_{k\in K} Y_{ijktv} \leq q_v Z_{ijtv} \qquad \forall i \in N, j \in N\backslash\{i\}, \forall v \in V, \forall t \in T \qquad (2)$$

$$\sum_{i\in N}\sum_{j\in N\backslash\{i\}} Z_{ijtv} \leq 1 \qquad \forall v \in V, \forall t \in T \qquad (3)$$

$$\sum_{j\in N\backslash\{i\}} Z_{jitv} - \sum_{j\in N\backslash\{i\}} Z_{ijtv} = 0 \qquad \forall i \in N, \forall v \in V, \forall t \in T \qquad (4)$$

$$\sum_{j\in N\backslash\{i\}} Z_{jitv} + \sum_{j\in N\backslash\{i\}} Z_{ijtv} = 2W_{it} \qquad \forall i \in N, \forall v \in V, \forall t \in T \qquad (5)$$

$$\sum_{i\in N\backslash\{0\}}\sum_{v\in V} Y_{oiktv} = \sum_{i\in N\backslash\{0\}}\sum_{v\in V} X_{iktv} \qquad \forall k \in K, \forall t \in T \qquad (6)$$

$$\sum_{j\in N\backslash\{i\}} Y_{jiktv} - \sum_{j\in N\backslash\{i\}} Y_{ijktv} = X_{iktv} \qquad \forall i \in N\backslash\{0\}, \forall k \in K, \forall v \in V, \forall t \in T \qquad (7)$$

$$X_{iktv} \geq 0 \qquad \forall i \in N\backslash\{0\}, \forall k \in K, \forall v \in V, \forall t \in T \qquad (8)$$

$$Y_{ijktv} \geq 0 \qquad \forall i \in N\backslash\{j\}, \forall j \in N\backslash\{i\}, \forall \in K, \forall v \in V, \forall t \in T \qquad (9)$$

$$W_{it} \in \{0,1\} \qquad \forall i \in N\backslash\{0\}, \forall t \in T \qquad (10)$$

$$Z_{ijtv} \in \{0,1\} \qquad \forall i \in N\backslash\{j\}, \forall j \in N\backslash\{i\}, \forall v \in V, \forall t \in T \qquad (11)$$

Constraints (2) guarantee that the vehicle capacity is not exceeded. Constraints (3) ensure that each customer is assigned to at most one vehicle. According to constraints (4), each vehicle must return to the depot by the end of planning period, constraints (5) are to ensure that each node is connected to two nodes. The commodity conservation flows at the depot and at the customers are represented by constraints (6) and (7), respectively. Finally, constraints (8-11) represent the types of decision variables.

3.2 Solution Methodology

To be able to solve medium and large-sized instances of the designated heterogeneous CVRP, two methods will be used; K-means clustering is used to partition the customers and aggregate them into clusters, and valid inequalities on routing are added to the mixed-integer programming model to make an approximation of the convex hull of the polyhedron of the problem. After solving the problem for the clusters, the model is disaggregated into the original locations to find the detailed tour for each vehicle.

3.2.1 K-means clustering

The number of customers in the target supply chain is assumed to be finite; hence, the number of clusters is also finite, and then the number of non-empty clusters *k* for *n* customers is a Stirling partition number that is given by:

$$\mathrm{S(n,k)} = \frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}\, i^n \qquad (12)$$

and the right side is zero when k > n.

A full enumeration is not possible especially for a large number of customers where the number of different combinations will be all the summations for equation (1) starting from a single cluster to the maximum number of

clusters. So, finding a global optimal partitioning solution is an NP-hard problem that is not practical to solve especially in dynamic supply chains such as in the case of FMCG.
In order to solve this problem, partitioning-based iterative clustering will be used to cluster the customers before performing the routing phase. K-means clustering is the most popular clustering method and is used to develop separated clusters, so that each customer is assigned to exactly one cluster. K-means clustering is an iterative relocation algorithm that minimizes or maximizes the value of a selected criterion or criteria until convergence. The most important advantage for the K-means clustering method is that it consumes less memory resources compared to other clustering methods such as hierarchical clustering analysis (Huang and Kim, 2006). However, there are some drawbacks with the traditional K-means algorithm, so in this work, the enhanced K-means algorithm proposed by Nananukul (2013) is elaborated to group customers into core clusters in a balanced way (He et al., 2009).

For a customer location (a_x, a_y) and a cluster's centroid (c_x, c_y), the Euclidean distance between them is given by:

$$d(a, C) = \sqrt{(a_x - c_x)^2 + (a_y - c_y)^2} \tag{13}$$

The difference between a customer location that belongs to cluster α and another cluster β can be given by:

$$dif(a, \beta) = d(a, \beta) - d(a, \alpha) \tag{14}$$

Figure 1 illustrates the algorithm for this procedure.

Step 0 Initialize the number of clusters.
Step 1 Assign customer *i* arbitrarily to the centroid of each cluster.
Step 2 Assign the remaining customers to the nearest clusters.
Step 3 Calculate summations of squared errors between customers' locations and cluster's centroid.
Step 4 If summation of squared errors is < 10^{-5} Go to step 5, otherwise repeat steps 1 to 3.
Step 5 Calculate the difference in the number of customers between every two clusters α and β, if it is < θ, STOP, otherwise Go to step 6.
Step 6 For two clusters α and β where difference in the number of customers > θ, *calculate diff*(a, β) for all $a \epsilon \alpha$. Sort the values in ascending order and move the first customer to cluster β, then repeat step 5.

Figure 1. Algorithm for the K-means clustering

In that way, it is possible to find the customers on the borders of different clusters and transfer them to the cluster with fewer customers, so that the clusters are balanced in terms of the number of customers in each cluster, the difference in the number of customers between any two clusters has a threshold θ. After performing the clustering, the MIP model presented in section 3.1 is solved for clusters instead of customers to assign vehicles to clusters.

3.2.2 Valid inequalities

Figure 2 shows cases of infeasibility due to path discontinuity or a route that does not contain the depot. Since the model includes a heterogeneous fleet of vehicles, then the known valid inequalities of the CVRP can be extended to strengthen the mathematical formulation. The following subtour elimination valid inequalities can be added to the model; see Cornuejols and Harche (1993) and Yaman (2006):

$$\mathrm{Z}(\mathrm{S}: \mathrm{S} \subset \mathrm{N}) \leq |\mathrm{S}| - \left\lceil \frac{d(S)}{q} \right\rceil \tag{14}$$

$$\sum_{v \in V: d(S) \leq q_v} Z(S: S \subset \mathrm{N}) + \frac{|S| - 1}{|S| - 2} \sum_{v \in V: d(S) > q_v} Z(S: S) \leq |\mathrm{S}| - 1 \tag{15}$$

where $|S| \geq 3$ *and* $d(S) \leq q$

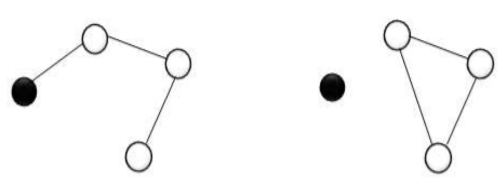

Figure 2.
(a)Path discontinuity (b) Route does not include depot

Hence, it is guaranteed that if the solver returned a solution x where $x \notin S$, then the corresponding graph must be disconnected from the original graph, and these subtours will be eliminated.

4. Computational Results

In this work, the data instances by Archetti et al. (2011) are used to test the proposed solution approach. The original data instances contain three classes with 14, 50, and 100 nodes and 6 time periods. Since the solution approach proposed in this work is designed to enable solving relatively big problems, then only the instances with 100 customers were used to test the ability of the solution approach to find a solution in a reasonable computational time. The data was extended by using the same guidelines used by Armentano et al. (2011) to include two products, and three heterogeneous vehicles, see Mostafa and Eltawil (2015b). Instances were generated for a number of 40 variations.

Since the proposed VRP problem is an operational supply chain management problem, the maximum computational time to solve a single run was limited to 2 hours (7,200 seconds). Due to this limit on the computational time, the optimality gap cannot be set to zero, to strengthen the results; the same gap used by Ruokokoski et al. (2010) of 10^{-6} will be used.

The K-means algorithm was implemented in Matlab R2012a and the mixed-integer programming model was solved by using IBM ILOG CPLEX 12.2 on a DELL Intel® Core (TM) i7 2.89 GHz with 16 GB RAM running Windows 7 Professional operating system.

Figure 3 shows the results after performing the K-means clustering approach, the 100 customers were partitioned in three balanced clusters according to their geographical locations; the number of customers in the three clusters is 31, 26 and 43, respectively. Figure 4 shows the best solution for one of the instances, the route is case-based given the demand of each customer and the capacity of each vehicle.

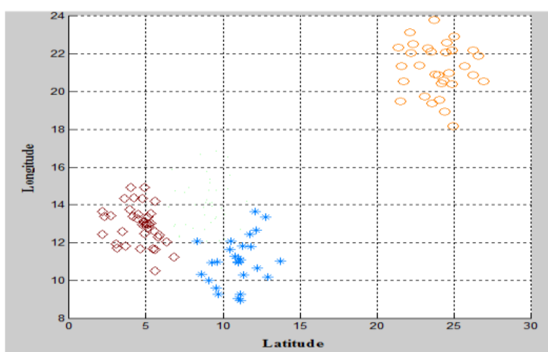

Figure 3. Results of K-means clustering

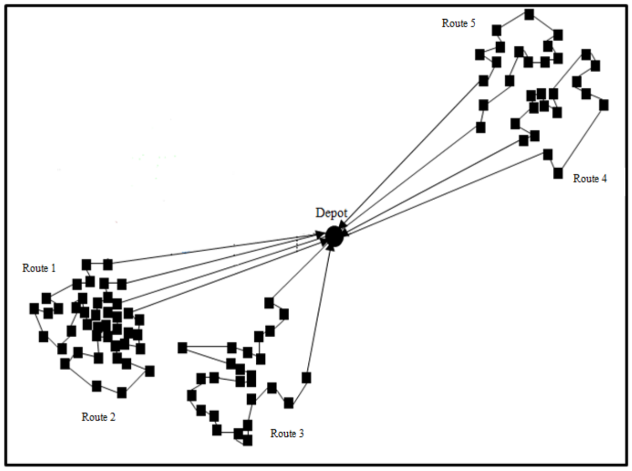

Figure 4 The best solution for one of the instances

It can be seen that two clusters are served by two vehicles for each cluster, and one cluster is served by a single vehicle. Since the propose approach is based on a cluster first-route second scheme, the clusters themselves are not changing, but the assignment of vehicles to clusters is varying depending on vehicle capacity and customer demand. Tables 1 and 2 show the effect of using the proposed approach in terms of total distance, vehicles' capacity utilization, and computational time. The first column of Table 1 shows the time period. Columns 2 and 3 show the average total distance travelled after solving the problem with the MIP model only and after using the proposed approach; Clustering and Valid Inequalities (CVI), respectively. Column 4 shows the deviation between the two methods. Columns 5 and 6 show the average percentage utilization for the vehicles by the two approaches, and column 7 shows the deviation between the two methods. Table 2 shows the average computational time for the 40 instances by using the MIP solution and after using the K-means clustering and valid inequalities approach.

Table 1: Comparison between the exact model and the proposed method

t	Av. Total distance			Av. Vehicle capacity utilization		
	MIP	**CVI**	**Dev.**	**MIP**	**CVI**	**Dev.**
1	352.85	364.17	0.032	95.51%	99.42%	0.041
2	360.48	366.36	0.016	92.37%	98.16%	0.063
3	368.53	390.88	0.061	95.40%	97.35%	0.020
4	355.18	381.43	0.074	94.66%	98.47%	0.040
5	373.72	392.50	0.050	97.48%	99.67%	0.022
6	349.64	353.49	0.011	94.18%	97.49%	0.035

Table 1 shows that the average decrease in the solution quality by using the proposed approach, CVI, does not exceed 5% compared to the exact solution. On the other hand, using the CVI approach could improve the vehicle utilization in all the time periods with an average of 4%. From Table 2, it was found that the exact model could not solve all the instances within the maximum computational limit, while the CVI approach was able to solve all the instances with a large reduction in the computational time with an average of 55%.

Table 2: Average CPU times (in seconds) for all the instances

I	Av. CPU time			i	Av. CPU time		
	MIP	**CVI**	**Dev.**		**MIP**	**CVI**	**Dev.**
1	415.02	255.43	0.385	21	310.26	257.33	0.171
2	380.71	301.86	0.207	22	502.42	278.61	0.445
3	392.84	308.41	0.215	23	538.16	364.15	0.323
4	850.27	413.50	0.514	24	7200	513.90	0.929
5	561.76	421.56	0.250	25	324.62	164.66	0.493

6	7200	562.37	0.922	26	7200	4116.10	0.428
7	853.75	365.15	0.572	27	425.63	237.16	0.443
8	534.70	246.76	0.539	28	251.05	107.30	0.573
9	643.85	275.16	0.573	29	597.34	363.72	0.391
10	463.85	391.07	0.157	30	1063.2	406.11	0.618
11	1263.40	497.56	0.606	31	7200	345.75	0.952
12	700.69	326.19	0.534	32	783.22	255.42	0.670
13	3513.24	546.23	0.845	33	573.91	318.93	0.444
14	800.63	369.43	0.539	34	980.63	367.43	0.625
15	623.81	279.43	0.552	35	3121.31	167.69	0.946
16	689.63	350.75	0.491	36	452.36	245.63	0.457
17	1601.70	298.46	0.814	37	707.64	345.12	0.512
18	7200	453.26	0.937	38	594.22	261.64	0.560
19	7200	3352.10	0.534	39	313.53	156.77	0.500
20	523.67	170.69	0.674	40	815.63	354.08	0.566

5. Conclusions

This paper considers large sized instances of the CVRP with heterogeneous vehicles. The problem was modeled as a MIP model, and then in order to solve the problem in a reasonable computational time, a two-step approach was proposed. In the first step, a balanced K-means clustering algorithm was developed to partition the customers into a number of balanced clusters. In the second step, two sets of valid inequalities are added to the model, and then the model is solved for the aggregated clusters. This solution is then disaggregated for the customers to build the routes within each cluster.

To test the proposed approach, data instances from literature with 100 customers were extended to test the efficiency of the proposed approach. From the computational results, it was found that the proposed clustering and cutting approach provides very good solutions with an average decrease in the quality not exceeding 5% compared to the exact solution, while improving the vehicle utilization with an average of 4% and reducing the computational time with an average of 55%.

Future work can be attempted towards integrating more supply chain management functions. The Production-Inventory-Distribution-Routing Problem (PIDRP) is a recent and complex problem that is hard to solve especially for large problems (Mostafa and Eltawil 2015a). With the promising results of the approach proposed in this work, it can be beneficial to use it to solve the IDRP. Another direction is to link the proposed approach with real-time traffic information by using technologies such as GPS and smart phone applications and the Internet of Things (IoT). This can be used to find real-time routing plans that would be very useful especially in case of tight time windows or emergency vehicles.

Acknowledgements

This research project was supported by the Egyptian ministry of higher education grant and the Japanese International Cooperation Agency (JICA) in the scope of the Egypt-Japan University of Science and Technology (E-JUST).

References

Abdelhalim, A., and Eltawil, A.B., A classification and review of recent models for solving the vehicle routing problem and a proposed new problem framework, *Proceedings of the IEEE International conference on Industrial Engineering and Management*, Hong Kong, December 10 – 13, 2012.

Agra, A., Andersson, H., Christiansen, M., and Wolsey, L., A maritime inventory routing problem: Discrete time formulations and valid inequalities, *Networks*, vol. 62, no. 4, pp. 297-314, 2013.

Archetti, C., Bertazzi, L., Paletta, G., and Speranza, M.G., Analysis of the maximum level policy in a production-distribution system. *Computes and Operations Research*, vol. 38, no. 12, pp. 1731-1746, 2011.

Armentano, V.A., Shiguemoto, A.L., and Lokketangen, A., Tabu search with path relinking for an integrated production-distribution problem, *Computers and Operations Research*, vol.38, no. 8, pp. 1199-1209, 2011.

Axsater, S., Aggregation of product data for hierarchical planning. *Operations Research*, vol. 29, pp. 744-756, 1981.

Balas, E., Cornuéjols, G., Kis, T., and Nannicini, G., Combining lift-and-project and reduce-and-split, *INFORMS Journal on Computing*, vol.25, no. 3, pp. 475-487, 2013.

Baldacci, R., Mingozzi, A., and Roberti, R., Recent exact algorithms for solving the vehicle routing problem under capacity and time window constraints, *European Journal of Operational Research*, vol. 218, pp. 1-6, 2012.

Barreto, S., Ferreira, C., Paixa, J., and Santos, B.S., Using clustering analysis in a capacitated location-routing problem, *European Journal of Operational Research*, vol. 179, pp. 968-977, 2007.

Bektas, T., and Laporte, G., The pollution-routing problem, *Transportation Research Part B: Methodological*, vol. 45, no. 8, pp. 1232-1250, 2011.

Belenguer, J.M., Martinez, M.C., and Mota, E., A lower bound for the split delivery vehicle routing problem, *Operations Research*, vol. 48, pp. 801-810, 2000.

Bertazzi, L., Bosco, A., Guerriero, F., and Laganà, D., A stochastic inventory routing problem with stock-out, *Transportation Research Par C: Emerging Technologies*, vol. 27, pp. 89-107, 2013.

Braekers, K., Ramaekersa, K., and Van Nieuwenhuyse, I., The vehicle routing problem: State of the art classification and review, *Computers and Industrial Engineering*, vol. 99, pp. 300-313, 2016.

Cagliano, R., Caniato, F., and Spina, G., E-business strategy: How companies are shaping their supply chain through the Internet, *International Journal of Operations and Production Management*, vol. 23, no. 10, pp. 1142-1162, 2003.

Campbell, A.M., and Wilson J.H., Forty years of periodic vehicle routing, *Networks*, vol. 63, no.3, pp. 2-15, 2014.

Cinar, D., Gakis, K., and Pardalos, P.M., A 2- phase constructive algorithm for cumulative vehicle routing problems with limited duration, *Expert Systems with applications*, vol. 56, pp. 48-58, 2016.

Coelho, L.C. and Laporte, G., A branch-and-cut algorithm for the multi-product multi-vehicle inventory-routing problem, *International Journal of Production Research*, vol. 51, no. 23-24, pp. 7156-7169, 2013.

Contardo, C., Cordeau, J.F., and Gendron, B., An exact algorithm based on cut-and-column generation for the capacitated location-routing problem, *INFORMS Journal on Computing*, vol. 26, no. 1, pp. 88-102, 2014.

Cornuejols, G., Valid inequalities for mixed integer linear programs, *Mathematical Programming Series B*, vol. 112, no. 1, pp. 3-44, 2008.

Cornuejols, G., and Harche, F., Polyhedral study of the capacitated vehicle routing problem, *Mathematical Programming*, vol. 60, no. 1, pp. 21-52, 1993.

Dantzig, G.B., and Ramser, J.H., The truck dispatching problem, *Management Science*, vol. 6, pp. 80-91, 1959.

Dondo, R., and Cerda, J., A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time window, *European Journal of Operational Research*, vol. 176, pp. 1478-1507, 2007.

Doulabi, S.H.H., and Seifi, A., Lower and upper bounds for location-arc routing problems with vehicle capacity constraints, *European Journal of Operational Research*, vol. 224, pp. 189-208, 2013.

Dror, M., Laporte, G., and Trudeau, P., Vehicle routing with split deliveries, *Discrete Applied Mathematics*, vol. 50, no. 3, pp. 239-254, 1994.

Erdogan, S., and Miller-Hooks, E., A Green Vehicle Routing Problem, *Transportation Research Part E*, vol. 48, pp. 100-114, 2012.

Ernst, R., and Cohen, M.A., Operations related groups (ORGs): a clustering procedure for production/inventory systems, *Journal of Operations Management*, vol. 9, pp. 574-598, 1990.

Ferrandez, S.M., Harbison, T., Weber, T., Sturges, R., and Rich, R., Optimization of a truck-drone in tandem delivery network using k-means and genetic algorithm, *Journal of Industrial Engineering and Managment*, vol. 9, no. 2, pp. 374-388, 2016.

Frohlich, M.T., and Westbrook, R., Arcs of integration: an international study of supply chain strategies, Journal of Operations Management, vol. 19, no. 2, pp. 185-200, 2001.

Garey, M.R., and Johnson, D.S., *A Guide to the Theory of NP-Completeness*, W. H. Freeman and company, New York, 1979.

Guerrero, W.J., Prodhon, C., Velasco, N., and Amaya, C.A., Hybrid heuristic for the inventory location-routing problem with deterministic demand, *International Journal of Production Economy*, vol. 146, no. 1, pp. 359-370, 2013.

Guerrero, W.J., Prodhon, C., Velasco, N., and Amaya, C.A., A relax-and-price heuristic for the inventory-location-routing problem, *International Transactions in Operational Research*, vol. 22, no. 1, pp. 129-148, 2015.

He, R., Xu, W., Sun, J., and Zu, B., Balanced K-Means Algorithm for Partitioning Areas in Large-Scale Vehicle Routing Problem, *Third International Symposium on Intelligent Information Technology Application*, pp. 87-90, 2009.

Huang, H., and Kim, K., Unsupervised clustering analysis of gene expression, *Chance*, vol. 19, no. 3, pp. 49-51, 2006.

Hwang, H.-S., Design of supply-chain logistics system considering service level, *Computers and Industrial Engineering*, vol. 43, no. 1-2, pp. 283-297, 2002.

Jin, M., Liu, K., and Bowden, R.O., A two-stage algorithm with valid inequalities for the split delivery vehicle routing problem, *International Journal of Production Economics*, vol. 105, no. 1, pp. 228-242, 2007.

Keeney, R.L., A method for districting among facilities, *Operations Research*, vol. 20, pp. 613-618, 1972.

Lin, C., Choy, K.L., Pang, G.K.H., and Ng, M.T.W., A data mining and optimization-based real-time mobile intelligent routing system for city logistics, *In Proceedings of the IEEE International Conference on Industrial and Information Systems*, pp. 156-161, Peradeniya, USA, December 17 – 20, 2013.

Melo, R.A., and Wolsey, L.A., MIP formulations and heuristics for two-level production-transportation problems, *Computers and Operations Research*, vol. 39, pp. 2776-2786, 2012.

Miranda, P.A., and Garrido, R.A., Incorporating inventory control decisions into a strategic distribution network design model with stochastic demand, *Transportation Research, Part E*, vol. 40, no. 3, pp. 183-207, 2004.

Mitchell, J.C., *Multi-item (s, S) inventory systems with service objective*, Report No. 26, University of North Carolina, Chapel Hill,1983.

Mitra, S., A parallel clustering technique for the vehicle routing problem with split deliveries and pickups, *Journal of the Operational Research Society*, vol. 59, no. 11, pp. 1532-1546, 2008.

Montoya-Torres, J.R., Franco, J.L., Isaza, S.N., Jiménez, H.F., and Herazo-Padilla, N., A literature review on the vehicle routing problem with multiple depots, *Computers and Industrial Engineering*, vol. 79, pp. 115-129, 2015.

Mostafa, N.A., and Eltawil, A.B., The production-inventory-distribution-routing problem: An integrated formulation and solution framework, *In Proceedings of the Industrial Engineering and Operations Management (IEOM)*, pp, 1-10, Dubai, UAE, March 3 – 5, 2015a.

Mostafa, N.A., and Eltawil, A.B., A Mixed-Integer Programming Model for the Production-Inventory-Distribution-Routing problem. *In proceedings of IEEE conference on Industrial Engineering and Engineering Management (IEEE IEEM)*, pp. 310-314, Singapore, December 6 – 9, 2015b.

Nananukul, N., Clustering model and algorithm for production inventory and distribution problem, *Applied Mathematical Modelling*, vol. 37, pp. 9846-9857, 2013.

Perboli, G., and Tadei, R., New families of valid inequalities for the two-echelon vehicle routing problem, *Electronic Notes in Discrete Mathematics*, vol. 36, pp. 639-646, 2010.

Pillac, V., Gendreau, M., Gueret, C., and Medaglia, A., A review of dynamic vehicle routing problems, *European Journal of Operational Research*, vol. 225, no. 1, pp. 1-11, 2013.

Praveen, V., Kousalya, K., and Prasanna Kumar, K.R., A nearest centroid classifier based clustering algorithm for solving vehicle routing problem, *In Proceedings of the 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, pp. 414-419, Chennai, India, February 27 – 28, 2016.

Rizk, N., Martel, A. an,d D'Amours, S., Synchronized production-distribution planning in a single-plant multi-destination network, *Journal of the Operational Research Society*, vol. 59, pp. 90-104, 2008.

Ruokokoski, M., Solyali, O., Cordeau, J.F., Jans, R., and Süral, H., Efficient formulations and a branch-and-cut algorithm for a production-routing problem, *GERAD Technical Report* G-2010-66, HEC Montréal, Canada, 2010.

Sahraeian, R., and Kaveh, P., Solving Capacitated P-Median Problem by Hybrid K-Means Clustering and Fixed Neighborhood Search algorithm, *In Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management (IEOM)*, pp. 1-6, Dhaka, Bangladesh, January 9 – 10, 2010.

Toth, P. and Vigo, D. eds., *Vehicle routing: Problems, methods, and applications*, 2nd edition, Vol. 18, MOS- SIAM series on optimization, Philadelphia, Pennsylvania : Society for Industrial and Applied Mathematics, 2014.

Wassan, N., and Nagy, G., Vehicle routing problem with deliveries and pickups: modelling issues and meta-heuristics solution approaches, *International Journal of Transportation*, vol. 2, no. 1, pp. 95-110, 2014.

Yaman, H., Formulations and valid inequalities for the heterogeneous vehicle routing problem, *Mathematical Programming Series A*, vol. 106, pp. 365–390, 2006.

Yucenur, G.N., and Demirel, N.C., A new geometric shape-based genetic clustering algorithm for the multi-depot vehicle routing problem, *Expert Systems with Applications*, vol. 38, pp. 11859-11865, 2011.

Biography

Noha Mostafa is a final-year PhD student in the department of Industrial Engineering and Systems management, of the Egypt-Japan university of Science and technology, Alexandria, Egypt. She received her B.Sc. in Industrial Engineering and Systems from Zagazig University, Egypt in 2007 and earned her M.Sc. in Industrial Engineering from Zagazig University in 2012. She was a visiting PhD student in Tokyo Institute of Technology, Japan in 2016. Her research interests include supply chain management, logistics management, operations research, and statistics.

Amr Eltawil is a professor and chairperson in the department of Industrial Engineering and Systems management, of the Egypt-Japan university of Science and technology, Egypt. The department is established in cooperation with the department of Industrial Engineering and Management of Tokyo Institute of Technology. His research interests are in the fields of applied operations research and discrete event simulation in supply chain design and management, and production planning and scheduling. Areas of interest are service systems planning and management especially container terminal related decisions, healthcare applications and call centres, as well as transportation and logistics and supply chain network design.