View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Procedia Computer Science
Volume 80, 2016, Pages 1887–1896
ICCS 2016. The International Conference on Computational Science

Integrated approach to assignment, scheduling and routing problems in a sales territory business plan

Laura Hervert-Escobar[1]*, Francisco López-Ramos[1,2], and Oscar A. Esquivel-Flores[1]

[1] Instituto Tecnológico y de Estudios Superiores de Monterrey. Monterrey, Nuevo León, México
[2] Pontificia Universidad Católica de Valparaíso. Valparaíso, Chile

Abstract
This paper considers a real life case study that determines the minimum number of sellers required to attend a set of customers located in a certain region taking into account the weekly schedule plan of the visits, as well as the optimal route. The problem is formulated as a combination of assignment, scheduling and routing problems. In the new formulation, case studies of small size subset of customers of the above type can be solved optimally. However, this subset of customers is not representative within the business plan of the company. To overcome this limitation, the problem is divided into three phases. A greedy algorithm is used in Phase I in order to identify a set of cost-effective feasible clusters of customers assigned to a seller. Phase II and III are then used to solve the problem of a weekly program for visiting the customers as well as to determine the route plan using MILP formulation. Several real life instances of different sizes have been solved demonstrating the efficiency of the proposed approach.

Keywords: Scheduling, Clusters, Customers, Routing, Visits

1 Introduction

Network models and integer programs are applicable to an enormous known variety of decision problems. In a real case, the cost efficient management decision is defined by a combination of different models. This paper considers a real life case study that determines the minimum number of sellers required to serve a set of customers located in a certain region together with the weekly schedule plan for visits and the optimal route. Therefore, the decision should consider the demand of the customers as well as the daily capacity of the sellers to fulfill the demand. Additionally, it is important to define per day the seller's route of visiting the customers.

*Corresponding author. E-mail:laura.hervert@itesm.mx, tel.: +52-181-1068-3479

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2016
© The Authors. Published by Elsevier B.V.

doi:10.1016/j.procs.2016.05.487

The solution method for the problem combines objectives and constraints of three classical approaches, the assignment of customers, the scheduling of visits, and the routing plan. Consider the location of the customers (such as points in the area or nodes of a network) with a given distance between every pair of points. We wish to find a cluster of customers using the nearest neighbor approach. Therefore, each cluster will represent a seller in the solution. This objective can be interpreted as the tightest cluster of m points. This is similar to the one facility version of the max-cover problem (for a network or discrete formulation [5], for planar models [10], for one facility [8], and for several facilities [14]) where we wish to find the location of several facilities which cover the maximum number of points within a given distance. After defining the clusters of customers, an important logistic problem to solve is the scheduling and routing problem.

In this context, the goal of the problem described above is to optimize the distribution process from depots to customers (routing design) in such a way that customer's demand of goods is satisfied without violating any problem-specific constraint. In the literature, these kind of logistic problems are known as Vehicle Routing Problems (VRP) and the objective regularly is the minimization of the complete distance traveled by the vehicles while servicing all the customers. The VRP is an interesting problem in operations research community due to its practical relevance and the difficulty to be solved exactly. Moreover, it is one of the most demanding NP-hard problems [7]. In reality, the task of finding the best set of vehicle tours by solving optimization models has a high computational cost, prohibitive for medium and large real applications.

Caceres et al. [3] present a survey on VRPs apply to real life problems. The authors call these VRPs as Rich (realistic) VRPs (RVRPs) and classify their variants according to the company decision levels and the routing elements involved. A classification that applies for this case study is Multi-Period/Periodic VRP with Multiple Visits/Split deliveries. In this classification, the clients are visited several times as vehicles may deliver a fraction of the customer's demand. Moreover, optimization is made over a set of days, considering a different frequency of visits to each client.

On the other hand, Bowerman et al. [2] classify the heuristic approaches to the VRP into five classes and suggest the development and analysis of approximate heuristic techniques capable of solving real-sized VRPs. The authors highlight the cluster-first/route-second heuristic, which first groups the nodes into clusters while assigning each cluster to a different vehicle, and then, and then find the vehicle tour by solving the corresponding traveling salesman problem (TSP). Xu et al. [15] provide a comprehensive and systematic description of the influential and important clustering algorithms rooted in statistics, computer science and machine learning, with emphasis on new advances. The authors depict the procedure of cluster analysis with four basic steps: a)feature selection or extraction, b) clustering algorithm for design or selection, c)cluster validation, and d)results interpretation. They illustrate applications of clustering methods for the TSP such as the divide and conquer one. This method gives the flexibility to hierarchically divide large problems into arbitrarily small clusters, depending on the desired tradeoff between accuracy and speed. Arpita et al. [11] classified the clustering algorithms into Hierarchically and Partition, they mention that the type of clustering algorithms used depends upon the application and the data set used in that field, there is no optimal solution for handling problems with large data sets of mixed and categorical attributes. Some of the clustering algorithms can be applied but their performance degrades as the size of data increases.

Effective scheduling systems aim at matching demand with capacity so that resources are better utilized and waiting times are minimized. Tuga and Emre [4] provide a comprehensive survey on appointment scheduling in outpatient services. The underlying problem applies to a wide variety of environments of outpatient scheduling, and is modeled using queuing system representing the unique set of conditions for the design of the patient appointment. The authors present a complete survey of problem definitions and formulations considering the nature of Decision-Making and Modeling of Clinic Environments. In addition, they mention a variety of performance criteria used in the literature to evaluate appointment systems, which are grouped as: a)cost-Based Measures, b)Time-Based Measures, c)Congestion Measures, d) Fairness Measures.

In recent years, an important and interesting variant of the VRP has been studied. This variant considers a stochastic version of the VRP where one or more parameters of the deterministic problem are considered as random variables, such as the number of customers to be served, the demands of the customers [12, 6], or the traveling times. Realistic routing problems include a scheduling part by incorporating traveling times between every pair of nodes, customer service times and maximum tour duration as additional problem data. A strategy to solve this VRP is to use time windows. Each customer has an associated time window defined by the earliest and the latest times where the service to the customer can start. The depot may also have a time window. Time windows can be hard or soft. In the hard time window, a vehicle is not allowed to arrive at the client/depot after the latest service start time; whereas, in the soft case, the time window can be violated (i.e., arriving later than the latest service time). Errico et al. [9] present a VRP with hard time windows and stochastic service times (VRPTW-ST). This VRP is modeled as a two-stage stochastic program, and two recourse policies are defined to recover operation feasibility when the first stage plan turns out to be infeasible. Finally, Pellegrini et al. [13] use a metaheuristic approach known as Ant Colony Optimization to tackle a VRP with multiple time windows/visits, considering a heterogeneous fleet and a maximum duration of the vehicle tours. The authors develop two variants of the method, The Ant Colony System and The *Max-Min* Ant Systems

In this work, a real business strategy for sales in different territories is modeled using the formulation of three classical problem: cluster, scheduling and VRP. Particularities of the modeling approach include scheduling constraints of visits spread over the week, service and traveling times; as well as time capacity to ensure the fulfillment of the clients demand.

The paper is organized as follows. A general mixed integer linear programming (MILP) formulation for the problem is presented in section 2. Section 3 describes a cluster greedy algorithm and two optimization models to cut the problem size and to reduce the solution time. In section 4, the models are tested for different scenarios. Finally, conclusions are presented in section 5.

2 Mathematical Formulation of the optimization model

Consider a set of customers $C = \{1, 2, ..., i, ..., j, ...N\}$ dispersed in a given region where their locations are given by coordinates (gx_j, gy_j). It is desired to design a business plan that includes the minimum number of sellers $S = \{1, 2, ..., s, ...S\}$ to attend the customers, in days $D = \{1, 2, 3, 4, 5, 6\}$ denoted by index t in the scheduling plan per week, and the optimal daily routing. Notation used in the mathematical formulation is described as follows:

PARAMETERS:

- gx_i, gy_i Location coordinates for customer i
- Dem_i Demand of customer i
- Cap Daily capacity of seller

- T_i Service time of customer i
- T^s Available time of seller s
- $Freq_i$ Number of times that a customer i is visited during the week. $Freq_i = \left\lceil \frac{Dem_i}{6*Cap} \right\rceil$
- M Big number
- d_{ij} Distance between two customers given by the following equation:

$$d_{ij} = -0.1677log\sqrt{(gx_i - gx_j)^2 + (gy_i - gy_j)^2} + 1.86846*log\sqrt{(gx_i - gx_j)^2 + (gy_i - gy_j)^2}$$

$\forall i, j \in N, i \neq j$

VARIABLES:

- Y_i^s Binary variable denoting whether customer i is assigned to seller s
- V_{it}^s Binary variable denoting whether seller s visits a customer i at day t
- $X_{ij}^{s,t}$ Binary variable denoting whether customer i is visited before customer j by seller s at day t
- e_i^{st} Continuous variable denoting the order in which customer i is visited in the route plan of seller s during day t

The mathematical formulation of the model is defined in the following equations:

$$min \sum_{s \in S} \sum_{i \in \mathbb{N}} 2^{s-1} Y_i^s + \sum_i \sum_j \sum_s \sum_t d_{i,j} \cdot X_{i,j}^{s,t} \tag{1}$$

subject to:

$$\sum_{s \in S} Y_i^s = 1, \ \ \forall i \tag{2}$$

$$V_{i,t}^s \leq Y_i^s, \ \ \forall i, t, s \tag{3}$$

$$X_{i,j}^{s,t} \leq V_{i,t}^s, \ \ \forall i, j, t, s, i \neq j \in N \tag{4}$$

$$X_{i,j}^{s,t} \leq V_{j,t}^s, \ \ \forall i, j, t, s, i \neq j \in N \tag{5}$$

$$\sum_i X_{i,j}^{s,t} = \sum_j X_{i,j}^{s,t}, \ \ \forall i, j, t, s, i \neq j \in N \tag{6}$$

$$\sum_i X_{i,j}^{s,t} + \sum_j X_{i,j}^{s,t} = 2V_{i,t}^s, \ \ \forall i, j, t, s, i \neq j \in N \tag{7}$$

$$\sum_i T_i \cdot V_{i,t}^s \leq T^s, \ \ \forall t, s \tag{8}$$

$$\sum_t \sum_s V_{i,t}^s = Freq_i, \ \ \forall i \tag{9}$$

$$V_{i,t}^s + V_{i,t+1}^s \leq 1, \ \ \forall i, s, t \leq 5 \ , \ Freq_i \leq 3 \tag{10}$$

$$e_i^{s,t} - e_j^{s,t} + MX_{i,j}^{s,t} \leq M - 1, \ \ \forall i, j, t, s, i \neq j \in N \tag{11}$$

$$e_i^{s,t} \leq \sum_j V_{j,t}^s, \quad \forall i, j, t, s, i \neq j \in N \tag{12}$$

The objective function (1) represents the sum of two goals, the minimization of the number of sellers required to service the customers and the minimization of the traveling distance to visit each customer for each routing plan.

As for constraints, (2) ensures that a customer is attended by only one seller. Next equation (3) guaranties that a customer is assigned to the seller that actually visits that customer. Equations (4) and (5) link the scheduling variables to the routing ones. Equations (6) and (7) are used for connectivity purposes. Equation (6) establishes that the number of incoming links to a client node must be equal to the number of outgoing links, whereas equation (7) sets for each client one arrival and one departure at the time (degree of the node is 2). Next equation (8) ensures that the available time of the seller is not compromised during the scheduling of visits to the customers assigned per day. The total visits per week is given by the frequency. The frequency is computed by dividing demand and capacity. In this way, equation (9) establishes the number of visits to carried out per customer according to the given frequency. Equation (10) avoids consecutive visits to those customers whose frequency is less than 4 visits per week. Finally, equations (11) and (12) allow to assign the proper order of visits to customers during the routing plan to avoid sub-tours. Equation (11) ensures that difference between the order of visits to two consecutive customers is one, whereas equation (12) limits the maximum order of visit to the customer.

3 Solution methods

The model combines assignment, scheduling and routing problem formulations. The above three models individually have been shown to be NP-complete [1]. Therefore, the time required to achieve an optimal solution of the whole model (problem) increases exponentially with the growth of the problem size. To overcome this difficulty, two approaches are proposed. In the first approach, the problem is divided into three sub-problems (or phases). The objective is to work with smaller problems at each phase. The first phase is treated as a cluster optimization problem where each cluster represents a seller and the clients are elements of the cluster. The second phase of the problem solves the scheduling problem for each seller. Finally, the third phase solves the routing problem for each active working day defined during the second phase. Another approach is to split the problem into two phases only. The first phase is the same as the one described in the previous approach (i.e. it performs the clustering), whereas the second phase combines the scheduling and routing of clients for each seller.

The first phase embraces the largest part of the problem size, thus, a greedy algorithm is proposed to determine the number of required sellers. The procedure is showed in Fig 1, starting with the activation of a seller. Then, the first customer is assigned to a seller. The assignment of customers must fulfill three conditions. The first condition is that the customer must be located within the covered area (this condition applies from the second assigned customer). The second condition is that the seller must have enough capacity to attend the demand of the customer to be assigned. Finally, the third condition ensures that the seller have enough time to attend the customer to be assigned and back to the depot. If after checking all customers, there is no candidate that fulfill all conditions, a new seller is activated. The procedure repeats until all customers have been assigned.

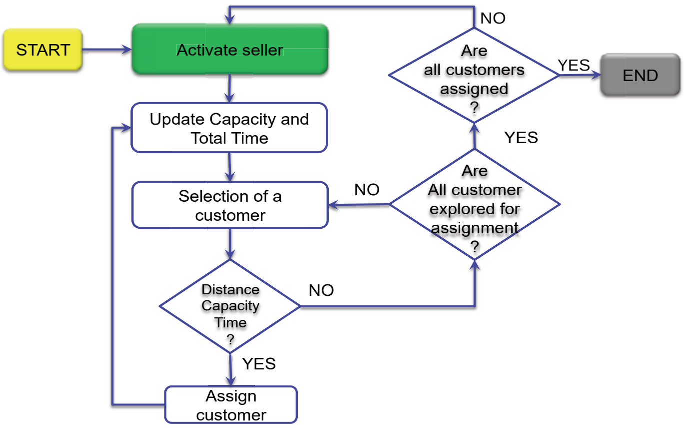

Figure 1: Greedy algorithm for the assignment phase.

As for the three-phase approach, the sequence of visits for each seller (cluster) is determined by solving a scheduling problem. Its objective function is given as follows (see (12)) :

$$min \sum_i \sum_t r_i \cdot V_i^t \tag{13}$$

The objective function (13) minimizes the radius of coverage (r_i) subject to constraints (8)–(10). This objective function helps to schedule per day the nearest customers. Finally, the routing is solved per day and seller in phase 3. The model of the third phase is based on the routing problem formulation. The objective function is given in (14).

$$min \sum_i \sum_j d_{i,j} \cdot X_{i,j} \tag{14}$$

The objective function (14) minimizes the total distance of the routes subject to constraints given by equations 4–7, and (11)–(12). When dealing with the two-phase approach (i.e. solving the scheduling and routing simultaneously), the optimization problem employs objective function (14) subject to constraints (3)–(12).

4 Results and discussion

To test the performance of the proposed models, several instances were tested. The data for each instance correspond to a real life case consisting of a soft-drinks manufacturer. Table 1 shows the nomenclature used to identify instances, where rows are used to describe the territory and columns denote the kind of seller required for the type of products. For each couple of territory and seller, the table provides the total number of customers to be assigned.

As shown in the table 1, the variety of size is good enough to prove whether the proposal is efficient for a business plan. The solving time is an important issue for the company due to the deadline to generate the business plan each week. Therefore, the results are given in terms of both objective functions as well as solving times. The greedy algorithm, which determines the total number of sellers needed to satisfy the customers demand, was implemented in C++ 9.0.21. The scheduling and routing models were implemented using AMPL to call the optimizer

Territory/Type of Seller	A	B	C	D	E	F	G
T1	48	37	163	15	970	145	10812
T2	33	12	463	4	1645	186	-
T3	26	18	405	9	1219	112	-
T4	59	40	448	17	1981	243	22475

Table 1: Customers to assign per territory and type of seller

CPLEX v.12.6.0. A time limit of 3600 sec is used as a stopping criteria when scheduling and routing are jointly solved. The cover area criteria for the greedy algorithm was set to 10 km.

The results are given in table 2 and figure 2. For each combination of territory and seller, the table provides the total number of customers per territory (# Custm), the total number of required sellers (# Sellers), the optimal solution provided by the scheduling solution from the three-phase approach (OF(Scheduling)) and the computational times of both approaches (T_{CPU}(2-p/3-p)) in seconds.

Territory	Type of Seller	# Custm	# Sellers	OF(Scheduling)	T_{CPU}(2-p/3-p)
T2	D	4	2	942	0.05/0.03
T3	D	9	3	1130	0.05/0.04
T2	B	12	1	4522	0.19/0.22
T1	D	15	5	1207	0.03/0.06
T4	D	17	6	1446	0.04/0.06
T3	B	18	1	6777	0.16/0.44
T3	AS	26	2	8472	0.19/0.4
T2	AS	33	5	5426	0.1/0.51
T1	B	37	1	13957	0.28/2.46
T4	B	40	1	15467	0.23/7
T1	A	48	4	7450	0.16/1.52
T4	A	59	7	6467	0.11/0.22
T3	F	112	3	28118	0.49/14.25
T1	F	145	4	25182	0.43/10.61
T1	C	163	5	26784	0.47/11.59
T2	F	186	5	27656	0.55/39.28
T4	F	243	6	28985	0.36/16.75
T3	C	405	11	38512	0.57/14.5
T4	C	448	12	29360	0.33/20.98
T2	C	463	11	38844	0.56/17.94
T1	E	970	4	91470	9.07/702.9
T3	E	1219	3	153303	29.85/1275.4
T2	E	1645	6	103299	4.26/611.46
T4	E	1981	12	62269	0.69/284.92

Table 2: The results of the greedy algorithm and the two approaches

It is observed that the total number of required sellers is not related to the size of the instance but to a combination of distance and demand of the customers. On the other hand, the objective function of the scheduling increases with the customers per territory. These results are expected since the objective function minimizes the distance of the radius of coverage, which is entirely related to the total number of customers to be visited. Regarding to computational time, the three-phase approach is faster than the two-phase one. Moreover, the three-phase approach achieves the same quality of the solution or even better. The computational performance improves with tight time windows and high node geographical density. Due to the use of the greedy algorithm, the critical size of the cluster-based MILP formulation significantly decreases and the hybrid approach becomes much more efficient.

Figure 2 shows a comparison of the objective functions for the routing solutions provided by the two approaches. The X-axis represents the relative gap between the values of the objective functions, whereas the Y-axis denotes the values of the objective functions. It is observed that the first approach gets better or equal solutions for the vast majority of instances, and that there is no pattern with respect to the size of the instances. However, the computation time is significantly different as mentioned above.

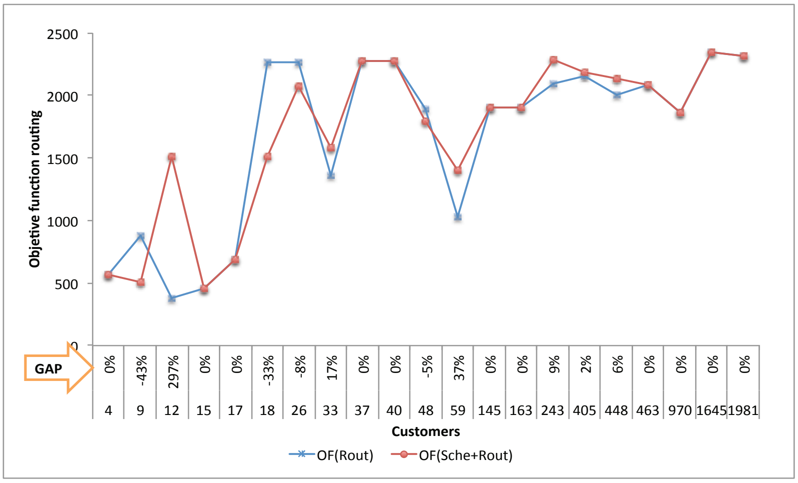

Figure 2: Routing results comparison

Next, figure 3 illustrates a clustering result for territory T3 and seller type C. On the figure, circles represent the locations of the customers in the territory according to the x-coordinates (Gx) and the y-coordinates (Gy). A different color is used to distinguish the sellers assigned to each customer as shown in the figure legend.

Finally, figure 4 shows a route example in a scheduling day with three sellers.

5 Conclusions

Assignment, scheduling and routing problems have attracted the attention of both researchers and practitioners for several decades due to the practical value of such models in decision-making contexts, the ever-present need and desire to incorporate increasingly realistic constraints and objectives into the models, the challenges associated with solving the models, and the ability of the basic formulations to represent important decision-making issues in business contexts. These four factors continue to this day and are likely to be present for years to come.

This work presents a two and three-phase approaches to solve a real life problem with different problem sizes. The first phase is used for both approaches and consists of a cluster-based algorithm that verifies time service and traveling time constraints in order to generate feasible clusters. The two-phase approach employs the cluster data to solve a smaller MILP model that determines the vehicle routes and their scheduling plan simultaneously, whereas the three-phase approach solves these two problems sequentially. First, the routing is performed,

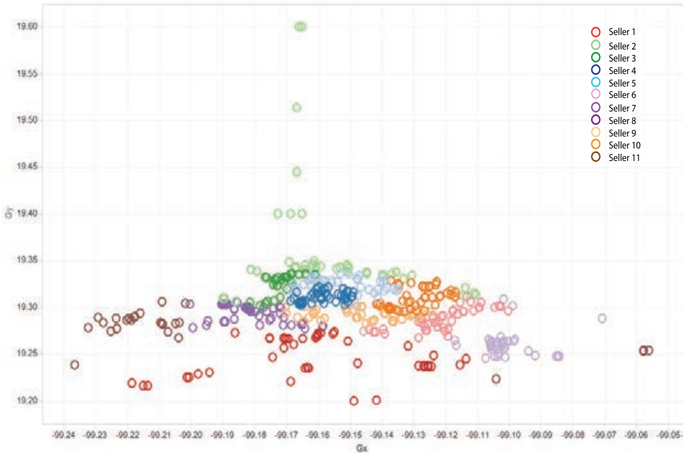

Figure 3: Example of customers assigned to a seller

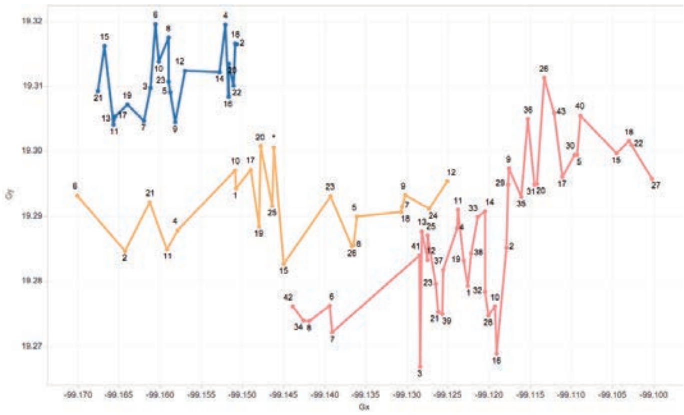

Figure 4: Route example result for three sellers one day visit

and then, the scheduling is set. The three-phase approach is capable of solving problems over 100 nodes at reasonable solving times. The results also indicate that the cluster-based optimization method proves to be quite successful on a variety of instances taken from real life cases. The computational time is a key factor when the solution approach is intended to be used as a making-decision. Therefore, a good but perhaps non-optimal solution is usually sufficient to obtain an efficient business plan.

The described approach allows tackling the uncertainties stemming from practical problems such as different sizes of territory and particular features of the demand such as the distance and the service time. Future research lines include the development of a metaheuristic for further improving the solution provided by the three-phase approach, as well as the addition of stochastic data to represent a raise/fall in the clients demand and the appearance/loss of clients. The latter clearly affects the service time, so alternative routes to the clients in a cluster should be sought. Moreover, this approach is better suited for parallel implementation for larger problems.

References

[1] Sara Baase and Allen Van Gelder. *Computer Algorithms: Introduction to Design and Analysis*. Addison Wesley, 1999.

[2] Robert L. Bowerman, Paul H. Calamai, and G. Brent Hall. The spacefilling curve with optimal partitioning heuristic for the vehicle routing problem. *European Journal of Operational Research*, 76(1):128 – 142, 1994.

[3] Jose Caceres-Cruz, Pol Arias, Daniel Guimarans, Daniel Riera, and Angel A. Juan. Rich vehicle routing problem: Survey. *ACM Comput. Surv.*, 47(2):32:1–32:28, December 2014.

[4] Tugba Cayirli and Emre Veral. Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, 12(4):519–549, 2003.

[5] Mark S. Daskin. *Network and Discrete Location: Models, Algorithms, and Applications*. John Wiley & Sons, Inc., New York, August 2013. ISBN: 978-0-470-90536-4.

[6] T.D. Dimitrakos and E.G. Kyriakidis. A single vehicle routing problem with pickups and deliveries, continuous random demands and predefined customer order. *European Journal of Operational Research*, 244(3):990 – 993, 2015.

[7] Rodolfo Dondo and Jaime Cerd. A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows. *European Journal of Operational Research*, 176(3):1478 – 1507, 2007.

[8] Zvi Drezner. Noteon a modified one-center model. *Management Science*, 27(7):848–851, 1981.

[9] F. Errico, G. Desaulniers, M. Gendreau, W. Rei, and L.-M. Rousseau. A priori optimization with recourse for the vehicle routing problem with hard time windows and stochastic service times. *European Journal of Operational Research*, 249(1):55 – 66, 2016.

[10] Mark Daskin John Current and David Schilling. Discrete network location models. In Horst W. Hamacher Zvi Drezner, editor, *Facility Location*, chapter 3, pages 81–118. Springer-Verlag Berlin Heidelberg, Oxford, 2002.

[11] A. Nagpal, A. Jatain, and D. Gaur. Review based on data clustering algorithms. In *Information Communication Technologies (ICT), 2013 IEEE Conference on*, pages 298–303, April 2013.

[12] D.G. Pandelis, E.G. Kyriakidis, and T.D. Dimitrakos. Single vehicle routing problems with a predefined customer sequence, compartmentalized load and stochastic demands. *European Journal of Operational Research*, 217(2):324 – 332, 2012.

[13] Paola Pellegrini, Daniela Favaretto, and Elena Moretti. *Knowledge-Based Intelligent Information and Engineering Systems: 11th International Conference, KES 2007, XVII Italian Workshop on Neural Networks, Vietri sul Mare, Italy, September 12-14, 2007. Proceedings, Part II*, chapter Multiple Ant Colony Optimization for a Rich Vehicle Routing Problem: A Case Study, pages 627–634. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[14] C.D.T. Watson-Gandy. Heuristic procedures for the m-partial cover problem on a plane. *European Journal of Operational Research*, 11(2):149 – 157, 1982.

[15] Rui Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, May 2005.