ISSN 2091-1521 (Print)
2091-153X (Online)
International Journal of Operational Research/Nepal - IJORN - 2019, Vol. 8, Issue 1

Spatio-Temporal Clustering for Optimizing Time-Sensitive Product Deliverie

Nicholas Chamansingh[1], Patrick Hosein[1], Elena Dyer[2]

[1] Department of Computer Science, The University of the West Indies, Trinidad
[2] Massy Technologies Ltd., Trinidad

Correspondence to: Nicholas Chamansingh, Email: nicholas.chamansingh@gmail.com

Abstract

We consider the problem of optimizing the near-periodic delivery of a product to customers with the additional constraint that there is an estimated deadline for delivery of the product. Delivery after the deadline is unacceptable while early delivery increases the frequency of deliveries which is also not desirable. While satisfying these customer constraints we wish to minimize the resources required by the delivery company to provide this service. In particular, we wish to minimize the total travel distance for delivery of products to customers since this reflects the associated costs of the service.

Keywords: Clustering, Optimal Routing, Rate Prediction, Delivery Optimization

1 Introduction and Contributions

Several papers have been published on delivery truck routing (e.g., see [11, 10, 7, 6, 8, 4, 2, 5]). However these focus on route optimization though some take into account constraints due to delivery of perishable items. We focus on a different type of problem. We assume that customers purchase a product that must be refilled (or replaced periodically). One example could be different types of gas products (oxygen, nitrogen, helium etc.) that must be refilled. In many cases the usage rate of the product could be used to estimate when next a refill is required. Given this information one can therefore better plan the routing of trucks to customers since the time of delivery constraint is more flexible than in the traditional approach (i.e., make a delivery when the customer makes a request). Since the model under investigation is different to those used in the past we have developed a unique solution approach using spatio-temporal clustering.

2 Related Work

The authors of [5] tackled the problem of customer satisfaction in terminal delivery in order to deal with a vehicle scheduling problem for perishable products. They were able to quantify customer satisfaction using information such as product freshness and time window. Their work had three main contributions, the first was based on a priority evaluation function and as a result they identified a service priority for each customer. The second was the construction of a multi-objective vehicle scheduling optimization model for perishable products considering the customer's satisfaction. Finally, their numerical experiments provided a clear indication of the effectiveness of the proposed method. That is, products with a certain shelf life range showed improved customer satisfaction. The results of the sensitivity analysis showed the adaptability of the proposed method in solving the perishable products terminal delivery.

On the other hand, [3] focused on solving the problem of Time Dependent Vehicle Routing and found solutions that minimized the number of routes to a given destination and the total travel time. Their solution involved a multi-ant colony system. Additionally, Abraham et al. [1] focused on a vehicle routing problem which found a set of trips, one for each vehicle. In order to solve the problem they employed a standard Genetic Algorithm. The vehicles would then pick up known quantities of perishable goods from a set of geographically dispersed suppliers and deliver them to the cargo terminal of an international airport in south India. Their objective was to serve a number of suppliers within a preferred time window at minimum cost while maintaining the capacity and total trip time constraints for each vehicle.

Similarly Toth et al. [9] considered the capacitated Vehicle Routing Problem. Their solution involved a cluster-first-route-second heuristic which utilized a new clustering method and can be used to solve problems with an asymmetric cost matrix. Their proposed approach exploits the information of the usually in-feasible vehicle routing problem and they also provide a lower bound on the solution.

Previous work done in vehicle routing primarily dealt with route optimization with some taking into account constraints relating to the delivery of perishable items and customer satisfaction. However, our approach assumes that customers purchase a product that must be refilled periodically. The usage rate of the product can be estimated and can therefore be used to better plan the routing of trucks to the customers. Hence the proposed model and the unique use of clustering makes our approach different from those discussed previously.

3 Problem Description

We assume that some consumable product is to be provided to customers. In particular, we focus on the delivery of refillable gas to customers. Each customer has one or more refillable gas tanks which must be refilled on a near-periodic basis. The refill period will depend on several factors such as the number of tanks, size of the tanks, usage rate of the gas, etc. Ideally the tanks should be refilled just when they are about to become empty since this will minimize the frequency of refills which minimizes the trips to the customer (saving on delivery resources) and this is also beneficial to the customer (since each visit can be inconvenient). However, because the gas exhaustion date has to be estimated then one must be conservative and make a visit prior to the estimated exhaustion date. The accuracy of the prediction and how fast a vehicle could be dispatched will determine how conservative one must be. Given these constraints, the problem is to minimize the travel distance of the delivery trucks (in order to save on fuel cost, truck maintenance costs and costs for the drivers). In the next section we describe the approach used to estimate the stochastic refill period of a customer. We then describe how deliveries are scheduled.

4 Predicting Time To Exhaustion

We assume that we are provided with historical data that contains, for the ith visit, the tank level v_i before refilling (as a fraction of capacity) and the date of the visit d_i (in number of days from the initial visit). In the case of multiple tanks we add the fraction left in the last tank with the total number of remaining full tanks and divide by the total number of tanks. For each consecutive pair of such readings we compute a smoothed estimate of the period T_i, as follows:

$$T_i = \alpha T_{i-1} + (1-\alpha)\frac{d_i - d_{i-1}}{1 - v_i}3 \qquad (1)$$

where $0 \leq \alpha \leq 1$ is the smoothing factor and T_i is the estimate of the period at the ith visit. The last fraction corresponds to the expected value of the last period (i.e the time it would have taken if the tanks were allowed to become empty). Finally we use this to estimate the next time for a refill $\hat{d}_{i+1}$ as

$$\hat{d}_{i+1} = d_i + T_i \qquad (2)$$

This estimate of the next time for refill is used in the clustering approach.

5 Spatio-Temporal Clustering

In order to reduce travel distance, once a truck visits a customer it should also visit as many nearby customers as possible. On the other hand, consider the case where all customers have the same number and size of tanks and the same gas usage rate. All customers served on a particular day will all need to be refilled at about the same time in the future. Therefore it is useful to cluster customers by expected exhaust date and serve all those with similar exhaust dates at about the same time and preferably do so near exhaustion. Hence we need to cluster in both space (serve all customers in the same neighborhood at the same time) and time (serve all customers who have similar exhaust dates together). We therefore cluster using the following 3 attributes, customer latitude, customer longitude and customer expected exhaust date. Note that, due to the nature of the data points, we use the K-Medoids algorithm for determining clusters. We determine the number of clusters K as follows. Assume we have N customers and the average time between refills (over all customers) is T days. This means, on average, N/T customers must be served per day. Suppose that we have M trucks then they must each serve about $N/(TM)$ customers per day and hence this should be the average size of our clusters. Therefore we choose $K = TM$. Once clusters are formed we order them by the average exhaust time of its members. We then choose clusters, starting from the cluster with the lowest average time before exhaustion, until all trucks are allocated. Note that if a cluster has too many members then we remove those furthest away until a suitable number is obtained. If a cluster has too few members then we add the closest ones first (ignoring those that have already been assigned to a cluster that is being served) until a suitable number is obtained. In Figure 1 we provide a simple illustrative example. In this example the dots are customer locations and the lighter colored dots have more immediate exhaustion dates and hence are more likely to be chosen.

On a daily basis all information is updated based on the prior day's deliveries and then K-Medoids clustering is used to determine which clusters of customers are to be served today.

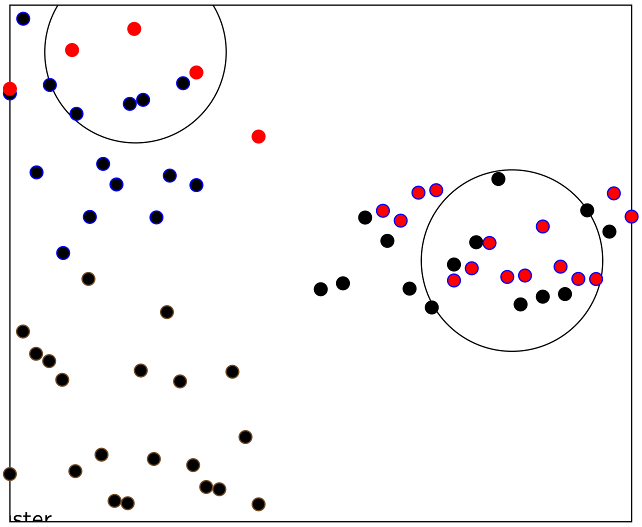

Figure 1: Customer Locations with urgent deliveries represented in red and purple

The attribute vector for each sample point (customer) consists of Longitude and Latitude values of the customer's location and the predicted exhaustion date of the customer (i.e., $\hat{d}$). Note that a scaling factor will have to be used when combining distance metrics for location and time. For example, if we have two customer vectors (x_1, y_1, d_1) and $x_2, y_2, d_2)$ then the distance between them is given by

$$D_{12} = |x_1 - x_2| + |y_1 - y_2| + \theta|d_1 - d_2| \tag{3}$$

where we have used the Manhattan Norm for the location separation and the factor θ is used to scale the time metric depending on the specific units used. If we denote the members of a particular cluster by $\mathcal{C}$ then the priority metric we use in deciding which clusters to serve is given by

$$P_{\mathcal{C}} = \frac{1}{|\mathcal{C}|} \sum_{j \in \mathcal{C}} \hat{d}_j \tag{4}$$

where smaller values of the metric (i.e., closer exhaustion date) have higher priority.

6 Simulation Setup

In order to compare and evaluate the proposed method to the present mode of operation, simulations were conducted. Customers were generated for the simulations. The customers were based off a small dataset provided by a local company containing a sample of customers that consume refillable gas. Each customer would purchase gas at some interval depending on their usage. For the simulations customers were generated based on the following constraints:

- 1 relatively central location was chosen for the gas distribution center.
- 10 distributed locations were selected based on the location of the distribution center.

- A pairwise distance was calculated between each location and the distribution center.
- Each location was used as a center point and 10 customers were generated within that center point for a total of 100 customers.
- All customers had gas containers with the same volume.
- Customers had an average usage of 14 days but randomly varied from 7 days to 21 days.
- A refill can only be done between Monday and Friday.

The simulations were generated to cover the course of 2 years (730 days). Since refills were done at a maximum of 21 days we ran the simulations for 35 iterations (730/21). Simulations were done for both the present mode of operation and the proposed method.

6.1 Present Mode of Operation

In the present mode of operation a customer has to routinely monitor the gas levels of their container. When the level is at 30% the customer would contact the gas supplier to have gas delivered to them before the gas is depleted. Each iteration of the simulation resulted in a number of customers which would be scheduled for a delivery based on a day of the week.

Note that refills are only performed when requested so, for example, consider the top diagram in Figure 2. On Monday there may be requests in both City A and City B followed by additional requests for both cities on Tuesday. If known, the Tuesday requests for City A could have been fulfilled on Monday and hence save a trip back to City A on Tuesday. This is the crux of our approach. We predict whether nearby customers will need refills soon and go ahead and make those refills when near them. In the present approach deliveries will be made to both City A and City B on both Monday and Tuesday.

6.2 Proposed Method of Operation

For the proposed method the next refill date is predicted on a daily basis. This prediction is based on a 20% threshold value, that is, when the gas container is at 20% of its total capacity. As described for the present mode of operation, each simulation would generate a number of customers which would be scheduled for a delivery. Consider again the example in Figure 2. On Monday the truck only visits City A and refills all customers that need to be refilled. However, in addition, we predict which customers will need to be refilled on Tuesday and go ahead and refill them. On Tuesday the truck no longer needs to visit City A and visits City B where is provides refills for customers who need refills (those identified for Monday and for Tuesday). Therefore the total distance travelled by the truck using the proposed approach is half that travelled in the present approach.

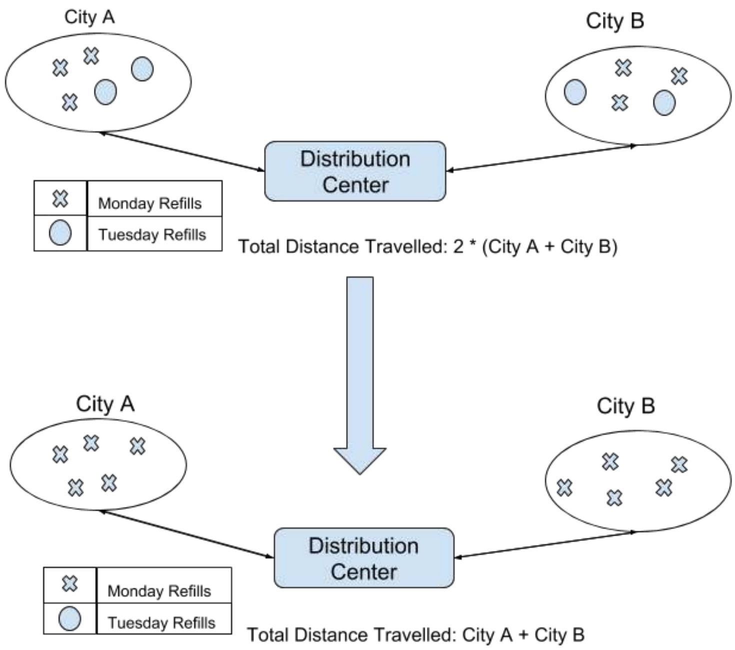

Figure 2: Distance Traveled Using the Present Method (Top) and Proposed Method (Below)

7 Data Description

The data generated for the simulations has two components. The first component is the location of the distribution center and the 10 locations that the gas supplier would have to visit. The data contained the following attributes:

- Location Name
- Latitude & Longitude
- Pairwise distance from the location of the distribution center

The second component of the data consisted of customer details. Each customer would have 1 of the 10 locations assigned to them. The customer data had the following attributes:

- Customer ID
- Number of days it took a customer to reach 30% of their gas supply
- Number of days it took a customer to reach 20% of their gas supply
- Customer Location
- The pairwise distance of the customer to the distribution center

8 Numerical Results

We use two metrics to determine the benefits of the proposed approach. The first is the fractional reduction in travel distance which indirectly represents the percentage reduction in delivery cost. The second is the fractional reduction in the number of visits to a customer which is also a desirable outcome since fewer visits means less customer inconvenience. Let us denote the number of simulated days by N and let $D_{pmo}(n)$ denote the distance travelled on day n with the present mode of operation and let $D_{pro}(n)$ denote the distance travelled with the proposed approach then the fractional decrease is denoted by

$$\rho_D = \sum_{n=1}^{N} \frac{D_{pmo}(n) - D_{pro}(n)}{D_{pmo}(n)} \tag{5}$$

Similarly if we denote the number of customer visits on day n by $V_{pmo}(n)$ and $V_{pro}(n)$ for the present and proposed methods of operation respectively then we have

$$\rho_V = \sum_{n=1}^{N} \frac{V_{pmo}(n) - V_{pro}(n)}{V_{pmo}(n)} \tag{6}$$

Note that since we target a lower depletion threshold in the proposed method than in the present method then the number of visits will be less.

For the simulations the total average percentage reduction in travel distance was 58.78% and the total average percentage frequency reduction in visits was 14.68%. In Figure 3 each bar represents the average percentage reduction in travel distance saved by the supplier with the proposed method. Figure 4 shows a sample of 20 customers and the average visit reduction with the proposed method.

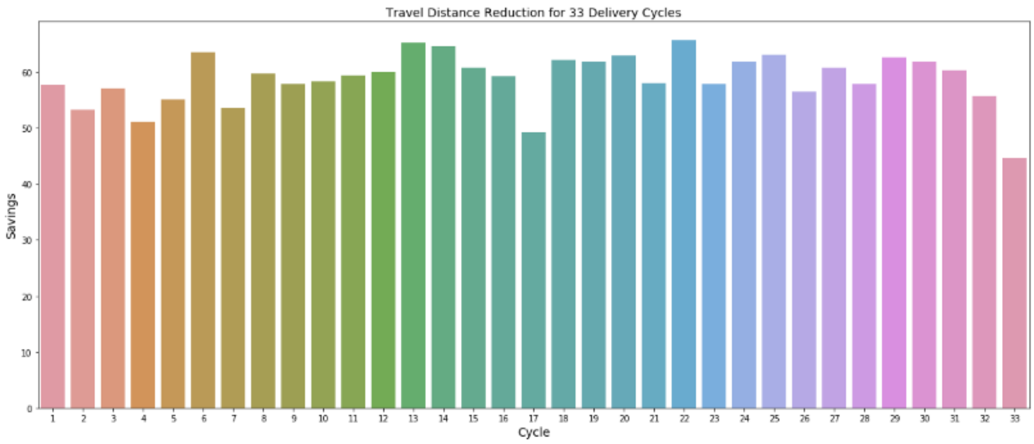

Figure 3: Percentage Reduction in Travel Distance versus Iteration

9 Discussion

In order to illustrate the benefits of the proposed approach we provide numerical results for simulated data but which is based on statistical characteristics of real data. In addition, we generate an artificial scenario with realistic assumptions, described in [6], so that we could do a sensitivity analysis of the proposed approach. We compare the proposed approach with the present mode of operation. In the present mode of operation the customer informs the delivery company when their gas supply has dropped to 30% and the company attempts to deliver gas to such customers on the following day. Therefore no exhaust prediction is performed and the scheduling does not take into account the exhaustion date. Note that a value of 30% has to be used to ensure that a delivery could be made to the customer before they run out of gas. For example, during busy periods, it may not be possible to dispatch soon after receipt of a customer request. However in the proposed method, the customer does not have to reach out to the supplier. In addition we predict when the gas supply will be at 20%. Because the predictions are done before hand, the company will be able plan their deliveries in advance. Using the scenario of busy periods, the company can increase the threshold value for the prediction of a customer's gas supply to relieve the overhead of the peak times. In addition, as a result of predicting when the customer's gas supply is at 20%, the company will make fewer trips to refill the customer. That is, we are refilling at 20% so the time between visits is the time taken to use 80%, whereas, in the present mode of operation, the time between visits is the time taken to use 70%. We compute the total distance traveled for both approaches. Since the major cost for the supplier is the distance traveled (which reflects fuel cost, maintenance cost and driver costs) then the percentage reduction in travel distance represents the percentage reduction in delivery cost, this was found to be 58.8%. The average reduction in visits to the customers was found to be 14.7%.

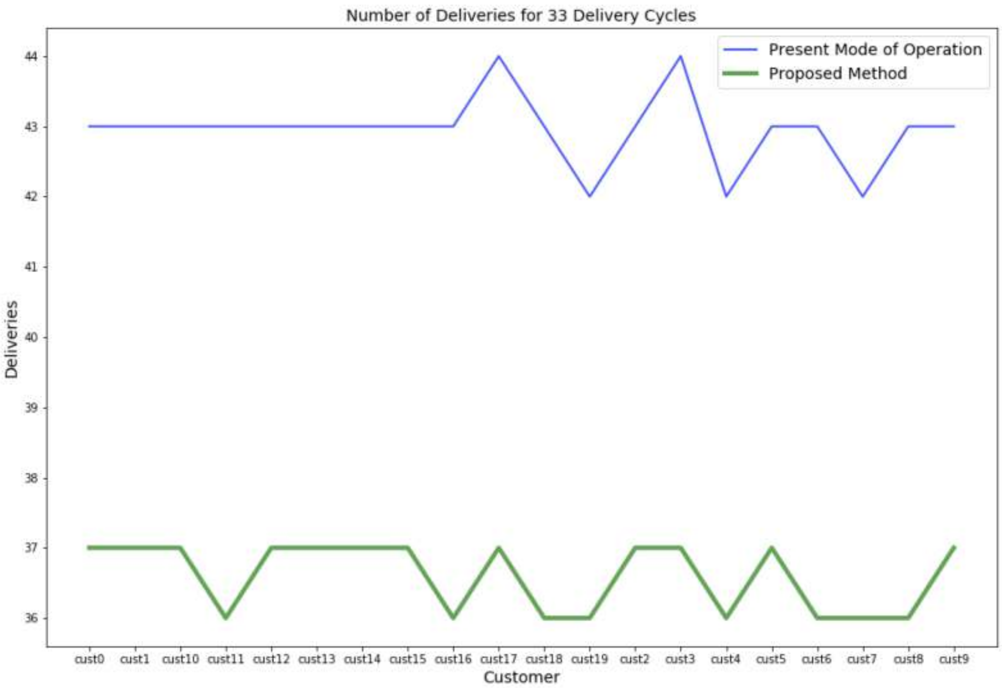

Figure 4: Total Average Percentage Frequency Reduction in Visits

10 Conclusion

We investigated a problem that has not been fully addressed in the past but one that is of interest to many companies, namely the delivery of some refillable product to a set of customers. We combine predictive analytics with clustering analytics and propose a simple, yet effective, solution to the problem. We show that the proposed approach can result in significant savings to the delivery company. In future work we plan to develop a commercial grade application for the company that will include additional features.

References

[1] A. K. Abraham, B. C. Jos, and G. S. Mangalathu. The pickup and delivery vehicle routing problem for perishable goods in air-cargo industry. *International Journal of Emerging Technology and Advanced Engineering*, 2(12):790–794, 2012.

[2] B. A. Clapp. Vehicle minimization for the multimodal pickup and delivery problem with time windows. Technical report, AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH GRADUATE SCHOOL OF ENGINEERING AND MANAGEMENT, 2013.

[3] A. V. Donati, R. Montemanni, N. Casagrande, A. E. Rizzoli, and L. M. Gambardella. Time dependent vehicle routing problem with a multi ant colony system. *European journal of operational research*, 185(3):1174–1191, 2008.

[4] A. Goel and V. Gruhn. A general vehicle routing problem. *European Journal of Operational Research*, 191(3):650–660, 2008.

[5] D. Juliandri, H. Mawengkang, and F. Bu'ulolo. Discrete optimization model for vehicle routing problem with scheduling side cosntraints. In *IOP Conference Series: Materials Science and Engineering*, volume 300, page 012024. IOP Publishing, 2018.

[6] C. Nguyen, M. Dessouky, and A. Toriello. Consolidation strategies for the delivery of perishable products. *Transportation Research Part E: Logistics and Transportation Review*, 69:108–121, 2014.

[7] D. Pisinger and S. Ropke. A general heuristic for vehicle routing problems. *Comput. Oper. Res.*, 34(8):2403–2435, Aug. 2007.

[8] N. I. Syahputri and H. Mawengkang. The time window vehicle routing problem considering closed route. In *Journal of Physics: Conference Series*, volume 930, page 012048. IOP Publishing, 2017.

[9] P. Toth and D. Vigo. A heuristic algorithm for the symmetric and asymmetric vehicle routing problems with backhauls. *European Journal of Operational Research*, 113(3):528–543, 1999.

[10] P. Toth and D. Vigo, editors. *The Vehicle Routing Problem*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.

[11] X. Wang, X. Sun, J. Dong, M. Wang, and J. Ruan. Optimizing terminal delivery of perishable products considering customer satisfaction. *Mathematical Problems in Engineering*, 2017, 2017.