Research Paper

International Journal of Research and Review
Vol. 9; Issue: 10; October 2022
Website: www.ijrrjournal.com
E-ISSN: 2349-9788; P-ISSN: 2454-2237

Solving Vehicle Routing Problem using Machine Learning based clustering and TSP Cluster Redistribution

Umesh Bhandari

Associate Development Architect, SAP Labs India Pvt. Ltd.

DOI: https://doi.org/10.52403/ijrr.20221041

ABSTRACT

Vehicle Routing Problem (VRP) is a well-known challenging nondeterministic polynomial-time hard (NP-hard) problem in logistics domain. Time window based VRP is an extension of the problem. The basic problem is to identify optimal set of routes for a fleet of vehicles to traverse to deliver to a given set of customers in a specified time duration. The objective is to minimize the costs of traversed routes. This paper proposes a machine learning and TSP based cluster redistribution approach to solve the time window based VRP. The proposed approach consists of three phases: machine learning based cluster creation, TSP based cluster routes and cluster re-distribution. The results demonstrate the efficacy and optimality of the proposed solution.

KEYWORDS: Vehicle Routing Problem (VRP), Time window based VRP, Machine learning

INTRODUCTION

In transportation management one of the important issues is cost-effective vehicle fleet management and scheduling of vehicles which are responsible for delivery/collection to a given no. of customers with pre-defined requirements. The transportation system manager needs to decide on which deliveries will be made by which vehicle, what will be the order of deliveries and total no. of vehicles used to minimize the overall cost.

Vehicle routing problem (VRP) is a very popular and challenging NP-hard problem. The VRP problem can be described as: Given a homogenous fleet of vehicle identify the order of deliveries to customer locations from a single depot so that every customer is served, and the costs of transportation i.e. total distance travelled by fleet routes is minimal. Each vehicle must start from the depot and the finally return to depot and has fixed capacity. Delivery to each customer will be made by only one vehicle. Time window based VRP (VRPTW) is an extension of the problem and focusses on deliveries to be made within a given time duration window. As VRPTW is also NP-hard it is computationally very expensive if not feasible to identify best set of routes for vehicles. VRPTW is explored and analyzed by many researchers in last few decades using exact approaches and lately using heuristic methodologies.

Machine Learning (ML) is a subfield of artificial intelligence, it involves computers learning how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data using different algorithms to carry out certain tasks and improve over the period. Machine learning provides different techniques and mechanisms to help us process data and provide meaningful information ex. clustering of data, identifying patterns and predictive analysis. Here in this research work ML based technique is leveraged to propose solution for VRPTW. In most research works the focus has been on identifying set of optimal routes for vehicle fleet but in this work besides it focus is on calculating optimal no. of vehicles required to fulfill deliveries along with delivery routes.

In this paper, the proposed solution consists of following 3 phases to solve the VRPTW problem

1. Machine learning based location cluster creation
2. TSP based cluster routes
3. Cluster re-distribution

PROPOSED ARCHITECTURE

VRPTW is a NP-hard problem so solving the problem is challenging for real world day to day applications which expect result for the input points in a certain amount of time. In past researchers have worked on different heuristics approaches to solve the problem. Every research solution had tried to solve the problem in a different way either using exact approaches or heuristic/meta-heuristics techniques or even genetic algorithms [1][2][3][4][5][6][7][8]. While every research suggested different way to solve the problem, but they have their own limitations and implementation implications in terms of time or effort for real-world applications.

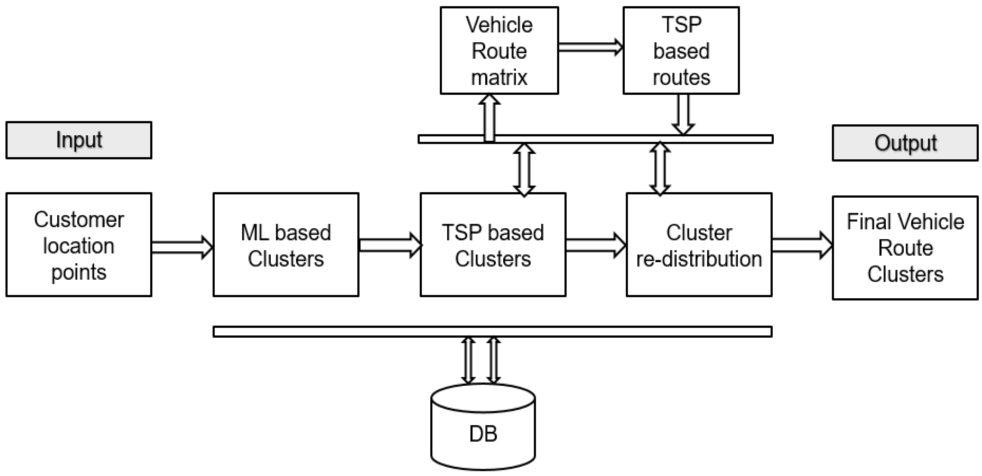

Figure – 1. Proposed Architecture

Fig. 1 shows an architecture of the proposed solution for the VRPTW problem for a given set of customer locations and generates the optimal no. of vehicles required along with their routes and sequence of deliveries to achieve customer deliveries within a given time window. This solution tries to leverage ML based technique to solve the problem.

The proposed architecture is explained below

- The input would be customer delivery location points and depot location
- Machine learning based component will process the input points and create clusters based on their spatial distance to each other. Cluster statistics like points inside each cluster, their route distance, route journey time and sequence of deliveries, etc. are stored in database and updated/used in each phase while processing.
- The ML clusters would be passed to next step for calculating actual routes between each location points inside a cluster and then TSP (Travelling salesman problem) algorithm would be used to calculate sequence of delivery location points in the route.
- The TSP based clusters interact with 'Vehicle Route Matrix' and 'TSP based routes' for calculating clusters.
- Each TSP based cluster will contain single route journey for a vehicle and if while calculation the route journey exceeds the time window then the cluster is further broken down into a

new separate cluster with additional delivery points.

- The TSP based clusters will be passed further to 'cluster re-distribution' phase as there may be many newly created clusters which are underutilized, or they may be better cluster combinations possible for existing clusters which are processed in this stage to create most optimal clusters.
- The final vehicle route clusters will be number of vehicles required to deliver at each location in given time window. Each cluster will denote the route journey for each vehicle and points inside the clusters will be served by one vehicle as per sequence of deliveries generated.

PROPOSED 3-PHASE SOLUTION AND EXPERIMENTAL ANALYSIS

Following are the 3 phases of solution in detail below along with experimental analysis:

Machine Learning based location cluster creation

Machine learning provides lot of algorithms for clustering, prediction, and analysis on data. In this section focus will be on machine learning based clustering algorithm to implement the first phase of proposed solution. ML Clustering is a technique for grouping on input data points based on certain features or attributes. K-means clustering is very simple, fast, and popular unsupervised clustering algorithm. Its simplicity makes it easy to understand and implement as most of the ML libraries provide direct support for the same. To brief about K-means algorithm, the 'means' in the k-means refers to averaging of data points i.e. finding centroid. A target number of clusters 'k' needs to be defined, which points to the number of centroids needed in the data. Basically, a centroid is the hypothetical or real location representing the center of the cluster. Every data point is assigned to each of the clusters by minimizing the in-cluster sum of squares. The algorithm starts with randomly selected centroids and the it performs iterative calculation to assigning data points to these centroids and optimize positions of centroids. It finally halts the iteration once the centroid much change between iterations.

As the main problem for solving VRPTW problem is first to calculate routes from each location point to every other location point and then these routes will be used in some combinatorial algorithm like TSP to identify sequence of deliveries which is optimal, this is a very time taking operation and even non-deterministic.

The basic idea behind the solution is instead of calculating actual routes from each customer location point to other, divide the whole set of location points into small clusters of points which are located close to each other based on spatial (or straight line) distance as calculating spatial distance is very simple and fast compared to calculating actual route.

This is based on the idea that location points which are located closely based on spatial distance are most likely closer by actual route distance. Then location points in these clusters will be used for route calculations only to other points within the cluster only, which would reduce overall processing time. Each cluster will depict the route journey for a single vehicle and points inside the cluster will be customer locations to be served by the vehicle within the given time in the resulted sequence of delivery.

For experimental analysis a realistic large city area has been chosen for service deliveries to 15,000 customer location points and time window for delivery is 2 hours for each vehicle.

Figure - 2. Input customer location points

Fig. 2 shows the 15,000 customer delivery location points denoted initially as black marker dots. These points will be used as input along with the depot location which is also located somewhere in between them. All the vehicle delivery route journey will start and end at the depot location. These input points will be passed to the 1st phase which is a ML based component which will process them and create clusters based on their spatial distance to each other.

ML algorithm K-means clustering will consider spatial distance between input location points and generate almost equal sized clusters of points which are closer to each other.

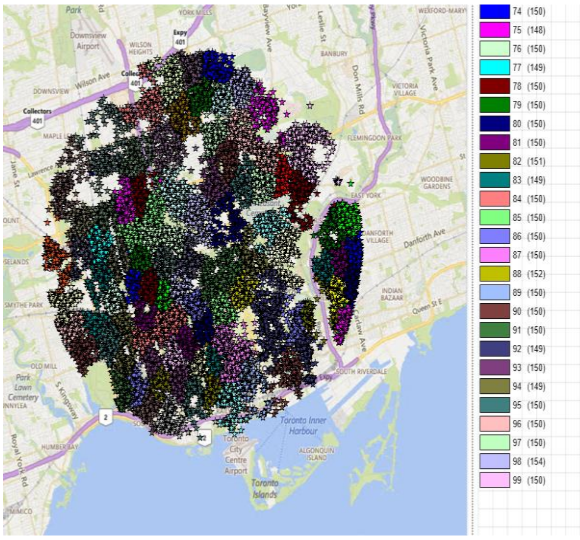

Figure - 3. Clusters generated in ML cluster phase

Fig. 3 shows the clusters generated in the ML cluster phase. Each cluster is denoted by different colored points inside them. As it can be seen location points located closer to each other are in same cluster and so have similar color. Only tricky part of using K-means clustering is choosing value for K initially as it is not known what actual no. of clusters would be formed beforehand. There are several methods available to derive value for K, here value of K is 100, so initial no. of clusters formed in ML cluster phase is 100 (Starting from 0 to 99) which can be seen in color coded count in Fig. 3 and each cluster has almost similar no. of points i.e. around 150 which shows ML based clustering as effective and efficient way to create clusters based on their spatial distance proximity.

TSP (Travelling Salesman Problem) based cluster routes

The 100 clusters created in ML based cluster phase will be passed to 2nd phase of TSP based cluster routes creation which will calculate actual routes between the points inside each cluster and create route journey for them.

Following steps will be performed in this stage:

- For each cluster calculate actual routing distance i.e. travel distance by road from each location point to every other point inside the cluster. There are several routing algorithms available for calculating the route distance between two points i.e. Dijkstra, A*, etc. [9].
- These route distance matrix from each point in cluster to every other point in cluster is feed to a TSP based algorithm ex. Simulated annealing [10] which uses meta-heuristic based approach.
- The TSP based algorithm will give route journey for each cluster which is most optimal order visiting each point in cluster within the given time window.
- In ML cluster phase as clusters were created based on spatial distance so each cluster point might not have been covered while calculating TSP using actual route within the given time window of 2 hrs. So, the points left out from journey in cluster have been removed them from existing cluster, a new cluster is created, and those points are assigned to it.
- As it can be seen in Fig. 4 the cluster count has increased to 122 in this phase compared to 100 in last phase as new clusters are created as well for points of cluster which were not covered in the time window.

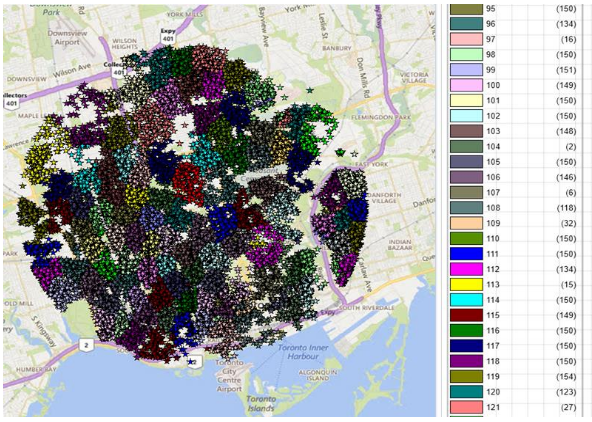

Figure - 4. Clusters after TSP cluster phase

Although the actual route calculated clusters are created now but on a closer look at Fig. 4, we can see that there are few clusters (the newly created ones) have very less location points ex. Cluster 104 has 2 points, cluster 107 has 6 points and cluster 113 has 15 points.

clusterId	count(*)	Time (min)
1	150	103.3
2	150	107.3333
3	152	112.9
4	152	116.1
5	124	109.2833
6	25	47.43333
7	147	119.85
8	3	22.48333
9	150	111.7333
10	148	110.6667
11	150	114.6333
12	137	118.7833
13	13	38.26667
14	134	111.3333
15	16	38.5

Figure - 5. Clusters statistics after TSP cluster phase

On further analyzing the statistics of the clusters it can be seen in Fig.5 that the clusters with less location points have total journey time very less than the maximum allowable time window of 2 hrs. (120 min) ex. Cluster 6,8,13 and 15 in Fig. 5. So, these cluster journeys are underutilized and needs to further optimize so that each cluster journey covers as much as possible location points in the given time window which will be the goal of the 3rd phase.

Cluster re-distribution

The 3rd phase of cluster re-distribution is the phase where the clusters from TSP based cluster phase are processed for further optimization and eventually reduce the number of clusters. In this phase focus is on 2 key points:

1. Look for merging of underutilized clusters so that maximum number of location points are covered in given time window
2. Try to optimize existing clusters as TSP route clusters were created based on 1st phase spatial distance clusters and there may be possibilities that other optimal cluster can be created by re-distributing locations points with nearby clusters

Following are the steps performed in the cluster re-distribution phase:

Assuming after 2nd TSP based distribution, all clusters will be less than max time size

1. Select clusters whose total time < 120 min or a threshold ex. 90-95% utilized
2. Form convex hull polygon [11] for each cluster using points in it and identify centroid. Then use this centroid to identify farthest points in cluster.
3. Sort points in clusters based on distance from hub
4. Starting from first cluster, select cluster(i) and iterate over all next (i+1) clusters to search points to merge till first cluster reaches threshold limit
5. Loop - If sum of cost of current cluster and iteration cluster is less than threshold limit then merge clusters
   - Run TSP on new cluster

   If
   - Time cost of cluster less than limit update cluster cost i.e. time and route distance

   Else
   - Split cluster and update cluster cost (case when cost may increase after merge due to additional distance between clusters)
   - Move to next cluster iteration
   - Else
   - Choose points from another cluster starting farthest from centroid to see if any point from them can be merged (Form convex hull polygon for each cluster using points in it and identify centroid. Then use this

centroid to identify farthest points in cluster)
- Loop points in [i+1] cluster
- Add point in current cluster and run TSP

If
- VRP time cost < limit then
- Add point in cluster and remove from [i+1] cluster
- Update cluster costs for both

Else
- Break from this loop of points and move to next cluster iteration

6. Repeat step 1

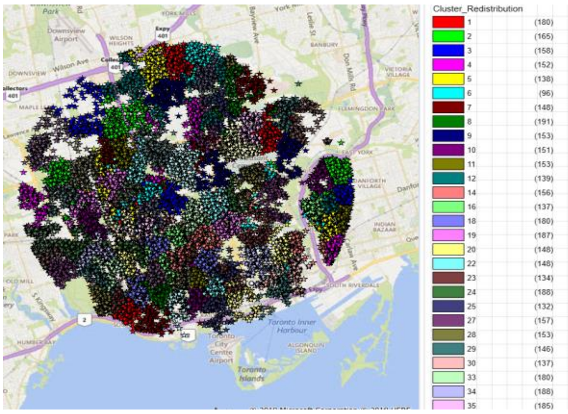

Figure - 6. Final clusters after cluster re-distribution phase

Fig. 6 shows the final clusters formed after cluster re-distribution stage. The total no. of clusters has reduced from 122 to 93 in this stage. In Fig. 6 on right side most clusters which were underutilized in previous stage are optimally filled with location point counts.

clusterId	count(*)	Time(min)
1	180	118.35
2	165	119.5667
3	158	119.6167
4	152	116.1
5	138	118.6
6	96	119.5167
7	148	118.2333
8	191	119.65
9	153	118.9667
10	151	119.5833
11	153	119.5667
12	139	119.5833

Figure - 7. Final clusters statistics after cluster re-distribution phase

Fig. 7 presents statistics of the clusters showcasing the fact that all the final clusters not only have optimal cluster point count but they are very much optimally utilized as well, it can be seen from the time covered by each cluster to cover all cluster points in a single journey is very close to the specified time window of 2 hrs.(120 min).

So, considering a single vehicle covers one cluster journey then they are covering optimal no. of customer delivery location points within the given time window. None of the cluster is underutilized as targeted clusters cover journey of locations points by utilizing at least 95% of time window which proves the efficacy of the proposed 3-phase solution.

RESULTS

All the analysis was conducted on a machine with configuration: 2.9 GHz, i7 Dual Core, 32 GB RAM. Average service time for each location: 15 sec was assumed (Results may vary based on it). Spectrum Spatial Routing[12] software and routing data was used to calculate actual route, route time and distance between location points. For TSP based route calculation to identify best route covering multiple location points opensource software JSPRIT[13] was used.

Following is the results summary
- Final no. of Clusters / No. of vehicles required to complete deliveries- 93
- Average journey time travel per cluster/vehicle- 117.92 min
- Average distance travel per cluster/vehicle- 30.28 miles
- Time Taken for processing/ complete calculation of clusters - ~11 hours

The cluster/vehicle numbers and average distance travelled and journey time per vehicle within the time window showcases the reasonable correctness and efficient output of the proposed approach. Another important aspect is the overall processing time i.e. time taken for complete calculation of final clusters which is around 11 hours considering the nature and complexity of problem i.e. scheduling and planning for optimal route journeys for 15,000 delivery location points given a 2 hour time window seems efficient.

CONCLUSION

The efficiency and correctness of the solution for VRPTW is critical for cost-effective vehicle fleet management and scheduling of vehicles. While the result's efficiency expectation is to have minimal no. of vehicles to fulfill all deliveries while the calculation time should be reasonable enough so that it can be performed even if required on daily basis with updated delivery locations. Over the years there have been different approaches explored ranging from exact approaches to meta-heuristics as well. Few of the approaches included clustering as part of solution [14][15][16] but they were different in many ways as they implemented clusters at either later stages of calculation or based on different parameter. In this paper, ML based clustering based on spatial distance and TSP based cluster re-distribution algorithm is proposed to solve the VRPTW problem and find optimal no. of vehicles required for fulfillment of deliveries. As analyzed and looking at the experimental observations the proposed solution generates results with optimal no. of vehicles within a reasonable amount of time. As the proposed solution is based on simple and generic algorithms it can be seamlessly implemented and integrated in almost any VRP solver implementation. The results also encourage for further study in proposed direction of exploring ML based techniques to solve and optimize the VRPTW problem.

Acknowledgement: All the experimental implementation related to this paper was done at Pitney Bowes Software India Pvt. Ltd [17]. All the resources i.e. hardware, software and data used for route calculation required for the experiments were provided by them.

Conflict of Interest: None

REFERENCES

1. R. Dondo and J. Cerdá, "A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows," *Eur. J. Oper. Res.*, vol. 176, no. 3, pp. 1478–1507, 2007, doi: 10.1016/j.ejor.2004.07.077.
2. M. Battarra, G. Erdoğan, and D. Vigo, "Exact algorithms for the clustered vehicle routing problem," *Oper. Res.*, vol. 62, no. 1, pp. 58–71, 2014, doi: 10.1287/opre.2013.1227.
3. R. Nallusamy, K. Duraiswamy, R. Dhanalaksmi, and P. Parthiban, "Optimization of Multiple Vehicle Routing Problems using Approximation Algorithms," no. March 2015, 2010, [Online]. Available: http://arxiv.org/abs/1001.4197.
4. J.-F. Cordeau, G. Desaulniers, J. Desrosiers, M. M. Solomon, and F. Soumis, "VRP with Time Windows," in *The Vehicle Routing Problem*, USA: Society for Industrial and Applied Mathematics, 2001, pp. 157–193.
5. B. Barán and M. Schaerer, "A multiobjective ant colony system for vehicle routing problem with time windows," *IASTED Int. Multi-Conference Appl. Informatics*, vol. 21, no. March, pp. 97–102, 2003.
6. M. Boujlil and S. Lissane Elhaq, "The vehicle routing problem with Time Window and Stochastic Demands(VRPTW-SD): Review," in *2020 IEEE 13th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA)*, 2020, pp. 1–6, doi: 10.1109/LOGISTIQUA49782.2020.9353927.
7. A. Ariyani, W. Mahmudy, and Y. P. Anggodo, "Hybrid Genetic Algorithms and Simulated Annealing for Multi-trip Vehicle Routing Problem with Time Windows," *Int. J. Electr. Comput. Eng.*, vol. 8, p. 4713, 2018, doi: 10.11591/ijece.v8i6.pp4713-4723.
8. D. Hou, H. Fan, X. Ren, P. Tian, and Y. Lv, "Time-Dependent Multi-Depot Heterogeneous Vehicle Routing Problem Considering Temporal–Spatial Distance," *Sustainability*, vol. 13, p. 4674, 2021, doi: 10.3390/su13094674.
9. "Shortest path problem - Wikipedia." https://en.wikipedia.org/wiki/Shortest_path_problem (accessed May 30, 2021).

10. “Simulated annealing - Wikipedia.” https://en.wikipedia.org/wiki/Simulated_annealing (accessed May 30, 2021).
11. “Convex hull - Wikipedia.” https://en.wikipedia.org/wiki/Convex_hull (accessed May 30, 2021).
12. “Spectrum Spatial Routing: Calculate travel times & distances.” https://www.precisely.com/product/precisely-spectrum-spatial/spectrum-spatial-routing (accessed May 30, 2021).
13. “GitHub - graphhopper/jsprit: jsprit is a java based, open source toolkit for solving rich vehicle routing problems.” https://github.com/graphhopper/jsprit (accessed May 30, 2021).
14. K. Shin and S. Han, “A Centroid-Based Heuristic Algorithm For The Capacitated Vehicle Routing Problem,” *Comput. Informatics*, vol. 30, pp. 721–732, 2011.
15. F. Oudouar and A. El Fellahi, “Solving the location-routing problems using clustering method,” *ACM Int. Conf. Proceeding Ser.*, vol. Part F1294, 2017, doi: 10.1145/3090354.3090472.
16. T. Vidal, M. Battarra, A. Subramanian, and G. Erdoǧan, “Hybrid metaheuristics for the Clustered Vehicle Routing Problem,” *Comput. Oper. Res.*, vol. 58, pp. 87–99, 2015, doi: 10.1016/j.cor.2014.10.019.
17. “Pitney Bowes IN | Digital Commerce, BI, Shipping & Mailing.” https://www.pitneybowes.com/in (accessed May 30, 2021).

How to cite this article: Umesh Bhandari. Solving vehicle routing problem using machine learning based clustering and TSP cluster redistribution. *International Journal of Research and Review*. 2022; 9(10): 344-351. DOI: *https://doi.org/10.52403/ijrr.20221041*

******