Cambridge Garbage Vehicle Routing using Integer Optimisation

15.093 OPTIMISATION METHODS

BRITTANY NGUYEN | YEN HANN YOO

MIT SLOAN SCHOOL OF MANAGEMENT

1.0 Motivation and Problem Statement

A fleet of refuse (garbage) vehicles are commonly employed to transport MSWs (municipal solid waste) from homes to waste treatment facilities. The three most prevalent types of treatment facilities are landfills, incinerators, and transfer stations. However, the routing of these vehicles is often not optimised. This has several negative implications such as cost for waste management businesses and emissions. Hence, the "vehicle routing problem" (colloquially referred to as VRP), within the context of optimising the routing of refuse vehicles from homes to landfill, was explored in this project for the city of Cambridge.

2.0 Data Gathering and Pre-Processing

Links to the datasets described in the following subsections can be found in Appendix B.

2.1 Housing Neighbourhoods Data

According to the City of Cambridge (2022), data from their "*Decennial Census Housing, Tenure, & Households 1980 – 2020*," indicates the presence of 13 neighbourhoods. For simplicity, only these 13 neighbourhoods of Cambridge were considered in this project instead of individual homes such that each of the 13 neighbourhoods are aggregated depictions of individual homes within their respective neighbourhoods. Columns pertaining to the following features were extracted from the data provided:

- Neighbourhood name
- Neighbourhood centroid coordinates (latitude and longitude)
- Number of households in 2020

2.2 Waste Treatment Facilities Data

Information on the location, status, and other features of various waste treatment facilities could be found from the State of Massachusetts website. For the purposes of this project, a single landfill was considered as the main treatment facility. It was assumed that all refuse vehicles serving Cambridge transport waste to the closest active landfill. By fixing the reference point for Cambridge to the MIT Sloan School of Management, the closest active landfill corresponded to *Taunton* at approximately 49.8km away. Hence, it was assumed that all refuse vehicles serving Cambridge begin from *Taunton* and eventually return to *Taunton* after serving their respective neighbourhoods.

2.3 Haversine Distances (d_{ij})

To approximate pairwise distances, the Haversine formulation was adopted. This is because the Haversine formula computes the spherical distance between two points and since the Earth better resembles a sphere than a flat plane, it was more appropriate to approximate distances using the Haversine formulation than using Euclidean distances:

$$d_{ij} = 2r \cdot arcsin\left(\sqrt{sin^2(\frac{\varphi_j - \varphi_i}{2}) + (cos\varphi_i)(cos\varphi_j)\, sin^2(\frac{\lambda_j - \lambda_i}{2})}\right)$$

Where:

- φ_i, φ_j: Latitude of point *i* and point *j* respectively
- λ_i, λ_j: Longitude of point *i* and point *j* respectively

Using this formulation, a 15x15 distance matrix was computed that represents the Haversine distances between neighbourhood-neighbourhood and landfill-neighbourhood pairings.

2.4 Cost Data (c_{ijk})

According to Hoare (2014), the operating cost of refuse vehicles can exceed $1.25 per mile (1.6km). Accounting for uncertainties and worst-case scenario, it would be reasonable to assume: $2 per mile. Standardising these values to a per km basis, the operating cost is **$1.25 per km of distance travelled per refuse vehicle**. Furthermore, the capacity of an individual refuse vehicle is 14 tons (12,700.59kg). With an aggregated depiction of homes as neighbourhood nodes, an aggregated depiction of refuse vehicles was also adopted. Each "aggregated vehicle" therefore has a corresponding capacity, C.

$$C = n_{IPA} \cdot 12700.59$$

Where n_{IPA} is a parameter defining the number of individual vehicles per aggregated vehicle. To reflect the aggregation of vehicles in the cost, the **cost matrix was also scaled up by n_{IPA}**. As there are 15x15 possible nodal pairings, the dimension of c_{ijk} is (15 x 15 x $n_{vehicles}$). Hence, the cost matrix c_{ijk} was computed as follows (assuming the operating cost across all *k* aggregated vehicles remains unchanged):

$$c_{ijk} = 1.25 \cdot n_{IPA} \cdot d_{ij} \quad \forall k = 1, \ldots, v$$

2.5 Emissions Data (ε_{ijk})

Azevedo, Jaramillo, & Tong (2015) state that a refuse vehicle travelling 25,000 miles annually emits 100 metric tonnes of eCO_2. This figure is equivalent to **2.5kg-eCO_2 per 1km travelled per refuse vehicle**. Therefore, assuming a homogeneous emissions factor across all $n_{vehicles}$, ε_{ijk} (emissions from travelling between node *i* and node *j* for vehicle *k*) can be computed as:

$$\varepsilon_{ijk} = 2.5 \cdot n_{IPA} \cdot d_{ij} \qquad \forall k = 1, \ldots, v$$

2.6 Waste Production Data (w_j)

According to the US Census Bureau (2020), there were 122,354,219 households in the US in 2020. According to USEPA (2022), 292.4 million tons of MSW was produced in 2018, an increase of 23.7 million tons from 2017. Assuming a constant increase of 23.7 million tons per year, 339.8 million tons of MSW were produced in 2020. Each household therefore produced 2.78 tons in 2020 (0.0076 tons per day per household). Standardising to kg, approximately 6.9kg-MSW per household was being produced each day in 2020. Therefore, the total waste produced per day per neighbourhood (w_j) is equivalent to:

$$w_j = 6.9 \cdot Number\ of\ households\ in\ Neighbourhood\ j$$

3.0 Formulation

A capacitated vehicle routing problem (CVRP) framework was adopted as refuse vehicles are typically constrained by how much waste they can carry. The landfill was appended as 2 additional nodes (1 and 15 specifically) to the set of 13 neighbourhood nodes (2 to 14), such that the landfill nodes have zero waste produced per day. The full formulation can be found in Appendix A.

Since n_{IPA} and $n_{vehicles}$ are parameters to be defined, the neighbourhood generating the largest amount of waste was *East Cambridge*, requiring four individual 14-ton capacitated refuse vehicles per day. Thus, the concept of "aggregated vehicles" each consisting of four individual 14-ton capacitated refuse vehicles was adopted. With n_{IPA} = 4, it was found that a minimum of $n_{vehicles}$ = 8 (i.e., 8 aggregated vehicles) was required to generate a feasible solution due to capacity constraints.

4.0 Results

A baseline model was also formulated for benchmarking (see Appendix C). A greedy search was utilised such that at each landfill or neighbourhood node (except node 15), we are looking for the cheapest unvisited neighbourhood node to visit next.

Table 1 | Cost minimisation comparison

n_{IPA} = 4 | $n_{VEHICLES}$ = 8	Baseline Model	Optimised Model
Total Operating Cost	$4084.57	$4049.52
	Total Savings	$35.05 (0.85%)

The routes taken by all vehicles was visualised as shown in Figure 1 (note the red dot represents the landfill translated from its true position for visualisation purposes), where each coloured arrow represents a vehicle *k*. Three vehicles serve multiple neighbourhoods (green, blue, and orange) whilst five vehicles serve a single neighbourhood due to the quantity of waste produced at these neighbourhoods that are sufficient to warrant a single “aggregated vehicle” to serve them.

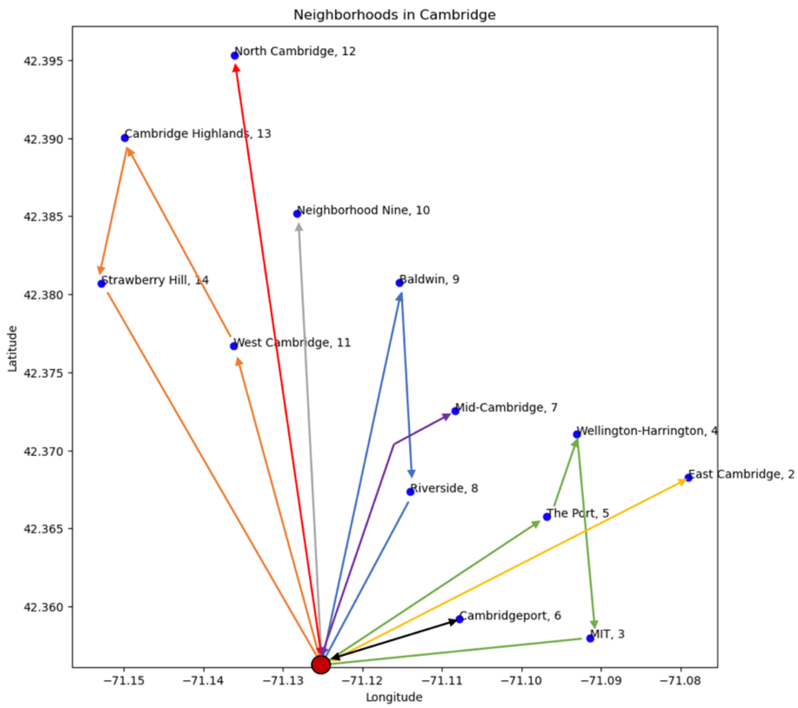

Figure 1 | Visualisation of vehicle routes (min. cost) for n_{IPA} = 4, $n_{vehicles}$ = 8

While minimising overall costs may not necessarily minimise emissions, the CVRP framework formulation can be easily modified to minimise emissions by substituting c_{ijk} for ε_{ijk}. Assuming all else unchanged, the same constraints and formulation can be adopted.

Table 2 | Emissions minimisation comparison

$\boldsymbol{n_{IPA} = 4 \mid n_{VEHICLES} = 8}$	Baseline Model	Optimised Model
Total Emissions [kg-eCO_2]	8157.24	8099.04
Total Reduction [kg-eCO_2] (%)		58.20 (0.85%)

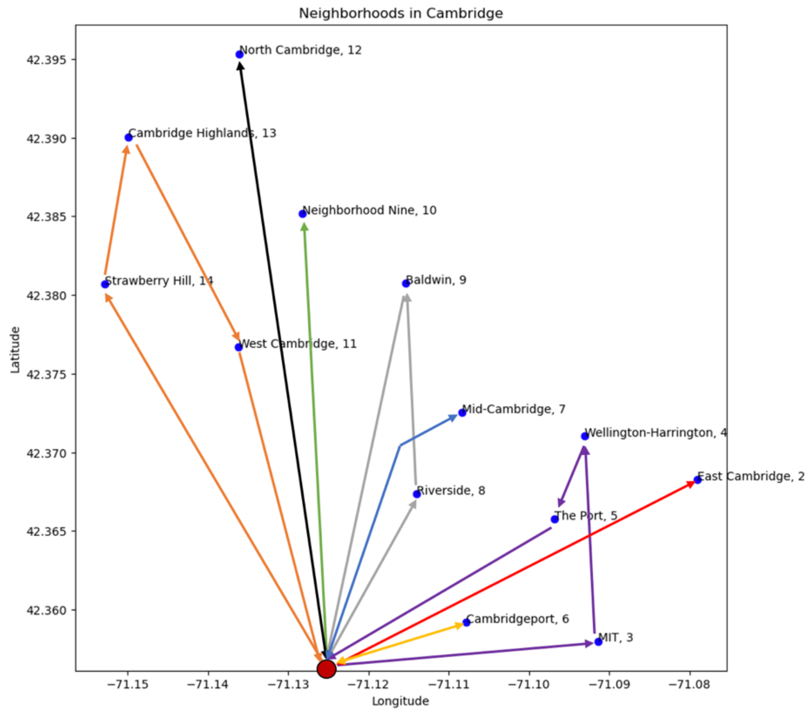

Figure 2 | Visualisation of vehicle routes (min. emissions) for n_{IPA} = 4, $n_{vehicles}$ = 8

Of note, the overall routes are identical for both minimising costs and emissions, suggesting that minimising one will minimise the other too. The reason for this is due to how the data was set up such that the emissions (ε_{ijk}) and cost (c_{ijk}) matrices were set up in a similar fashion: scaled up d_{ij} by a constant and n_{IPA}, as outlined in section 2.0. Thus, the costs and emissions are proportional. As one increases so too does the other and minimising one, will also minimise the other (i.e., no obvious trade-off relationship). Hence, the routes chosen would be the same when minimising both emissions and costs. An improvement has been suggested in 6.0.

5.0 Practical Implications

The first implication involves reducing the overall operating costs for waste management companies through the optimization formulation. Albeit the cost savings of $35.05 is relatively small, the formulation has proven to work for a small-scale problem in this project. Implicatively, for large-scale problems, the same framework can be expanded for more waste treatment facilities, nodes, and vehicles. As the problem scales in dimension, so too will the cost savings. Reducing these costs will allow companies to re-allocate funds towards other goals, such as serving underrepresented neighbourhoods or expanding the company's reach.

Another implication of the model is the potential reduction on emissions. The US Department of Energy (2020) reported that annually, refuse trucks have the 3rd highest average fuel consumption and the worst fuel economy, at approximately 2.5 MPG.

Therefore, the formulation can also reduce emissions to make waste disposal a "greener" process. As before, as the problem scales with dimension, so too will the total reduction in emissions.

6.0 Further Works

While the results discussed above have demonstrated savings and reductions, there are several potential areas to explore further. These include:

- *Restructure data and incorporate trade-Off parameter, λ*

  To accurately depict a trade-off relationship between costs and emissions, one would have to restructure how the data was set up, such that there is a non-proportional relationship between costs and emissions. One this is completed, a trade-off parameter $\lambda \in [0,1]$ can be incorporated into the objective function (as below) and a "sweet-spot" between cost and emissions can subsequently be found.

$$min \sum_{k=1}^{v} \sum_{i=1}^{15} \sum_{j=1}^{15} \lambda c_{ijk} x_{ijk} + (1-\lambda) \varepsilon_{ijk} x_{ijk}$$

- *Incorporating robustness*

  Costs and emissions are rarely ever fixed nor constant. Thus, it is imperative that robustness is incorporated into the formulation by introducing uncertainty sets for both costs and emissions. While efforts have been made to account for the worst-case scenario by scaling up to $2 per mile travelled, it would be more appropriate to account for uncertainty via the use of uncertainty sets and robust optimisation.

- *Incorporating multiple waste treatment facilities, vehicles, and nodes*

  While the formulation has proven to work for a small-scale problem in this project, the notion of aggregated neighbourhoods was adopted for simplicity. Thus, one should proceed further by disaggregating the neighbourhoods into individual homes (or at least, smaller clusters that leads to more "neighbourhood nodes"). Furthermore, multiple waste treatment facilities can also be considered and presumably, an assignment step would also follow to designate which vehicles go to which waste treatment facilities. This would be more reflective of real-world waste management.

- *Comparisons with column generation methods*

  Column generation methods can also be applied by pre-computing possible and feasible routes. Subsequently, results from column generation methods can be compared to the CVRP framework and the baseline model to see which performs better, whilst simultaneously taking computational efforts and time into account too.

7.0 References

Azevedo, I., Jaramillo, P. & Tong, F. (2015) *Comparison of life cycle greenhouse gases from natural gas pathways for medium and heavy-duty vehicles.* Available from: https://pubmed.ncbi.nlm.nih.gov/25938939/ [Accessed: 7th November 2022]

Hoare, D. (2014) *Cost per Mile – The Basic Formula*. Available from: https://businessecon.org/cost-per-mile-the-basic-formula/ [Accessed: 10th November 2022]

US Department of Energy (2020) *Average Annual Fuel Use by Vehicle Type.* Available from: https://afdc.energy.gov/data/10308 [Accessed: 4th November 2022]

US Department of Energy (2020) *Average Fuel Economy by Major Vehicle Category.* Available from: https://afdc.energy.gov/data/10310 [Accessed: 4th November 2022]

US Environmental Protection Agency (2022) *National Overview: Facts and Figures on Materials, Wastes, and Recycling*. Available from: https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/national-overview-facts-and-figures-materials [Accessed: 10th November 2022]

8.0 Appendix A: CVRP Formulation

Sets and Indices

- Landfill and neighbourhoods: $i = 1, 2, 3, \dots., 15$
- Landfill and neighbourhoods: $j = 1, 2, 3, \dots., 15$
- Vehicles: $k = 1,2,3 \dots., v$

Landfill nodes: 1 and 15, both represent the same landfill and $v = n_{vehicles}$ (for simplicity of writing)

Decision Variables

$$x_{ijk} = \begin{cases} 1, & if\ vehicle\ k\ serves\ i\ to\ j\ in\ sequence \\ 0, & otherwise \end{cases}$$

$$u_{ik} = Cumulative\ total\ waste\ collected\ by\ vehicle\ k, up\ to\ and\ including\ node\ i$$

Objective Function

$$min \sum_{k=1}^{v} \sum_{i=1}^{15} \sum_{j=1}^{15} c_{ijk} x_{ijk}$$

Note: If the goal is to minimise emissions, c_{ijk} can be substituted for ε_{ijk}

Constraints

1) No travel from a node to itself

$$x_{i,i,k} = 0 \quad \forall i = 1, \dots, 15, \forall k = 1, \dots, v$$

2) No direct travel between landfill nodes (nodes 1 and 15)

$$x_{1,15,k} = 0 \quad \forall k = 1, \dots, v$$

3) No outflow travel from landfill (node 15); all routes terminate at landfill node 15

$$x_{15,j,k} = 0 \quad \forall j = 1, \dots, 15, \forall k = 1, \dots, v$$

4) No inflow travel into node 1; all routes begin and depart from landfill node 1

$$x_{i,1,k} = 0 \quad \forall i = 1, \dots, 15, \forall k = 1, \dots, v$$

5) Each vehicle departing from landfill node 1 goes to exactly 1 neighborhood

$$\sum_{j=2}^{14} x_{1,j,k} = 1 \quad \forall k = 1, \dots, v$$

6) Each neighborhood j is visited exactly once (node 15 can be visited > 1 time)

$$\sum_{k=1}^{v} \sum_{i=1}^{14} x_{i,j,k} = 1 \quad \forall j = 2, \dots, 14$$

7) Each vehicle must return to landfill node 15

$$\sum_{i=2}^{14} x_{i,15,k} = 1 \quad \forall k = 1, \dots, v$$

8) Vehicle capacity cannot be exceeded

$$\sum_{i=1}^{14} \sum_{j=2}^{15} w_j x_{i,j,k} \leq C \quad \forall k = 1, \dots, v$$

9) Miller-Tucker-Zemlin (MTZ) subtour elimination

$$u_{j,k} \geq u_{i,k} + w_j - C\left(1 - x_{i,j,k}\right) \;\; \forall i = 1, \dots, 14 \;\; \forall j = 2, \dots, 15 \;\; \forall k = 1, \dots, v$$
$$w_i \leq u_{i,k} \leq c \quad \forall i = 2, \dots, 15 \;\; \forall k = 1, \dots, v$$
$$u_{1,k} = 0 \quad \forall k = 1, \dots, v$$

10) Each vehicle departs from the same node in which it entered

$$\sum_{j=1}^{14} x_{j,i,k} = \sum_{j=2}^{15} x_{i,j,k} \quad \forall i = 2, \dots, 14, \qquad \forall k = 1, \dots, v$$

11) Binary decision variable $x_{i,j,k}$

$$x_{i,j,k} \in \{0,1\} \qquad \forall i = 1, \dots, 15 \quad \forall j = 1, \dots, 15 \quad \forall k = 1, \dots, v$$

9.0 Appendix B: Data

Links to datasets: [Datasets]

10.0 Appendix C: Baseline Model Algorithm (Greedy Search)

Algorithmically, for any vehicle *k*:

1. From landfill node 1, travel to the cheapest unvisited neighbourhood node *j*

2. Sum the waste collected and set neighbourhood node *j* as node *i* (destination node from (1) is now set as the new origin node). From this neighbourhood node *i*, find the cheapest unvisited neighbourhood node *j*.

3. Check if visiting neighbourhood node *j* violates capacity constraints. If violated, set penultimate node on vehicle route as neighbourhood node *i* and return to terminal landfill node 15. If not violated, visit neighborhood node *j* and repeat (2) and (3) until capacity constraints violated or all neighborhood nodes visited once.

4. Compute total cost of operating vehicle *k*

Once (1) to (4) have been repeated across all vehicles *k*, the total baseline cost can be computed by cumulating the operating costs across all $n_{vehicles}$.