We wish to model the distribution of a product to a number of customers. Specifically we consider

i) a set of locations $\mathcal{N} = \{0, ..., n\}$, where location 0 is the starting point, location n is the end point and $\mathcal{B} = \mathcal{N} \backslash \{0, n\}$ is the set of customers;

ii) product demands $d_i \geq 0$ for all customers $i \in \mathcal{B}$;

iii) a delivery vehicle with capacity K;

iv) given distances between all pair of locations c_{ij};

v) a maximum total distance L the truck can travel.

We wish to establish a delivery route such that

i) the route starts at location 0 and ends at location n;

ii) the capacity of the truck is never exceeded;

iii) the maximum total distance of the truck is never exceeded;

iv) not all customers must be visited, but if a customer is visited *all* of their demand must be satisfied;

v) the total number of products delivered is maximized.

Above we have used the following assumptions

i) the distribution cost correlates only to distance traveled and

ii) the income from deliveries correlate only to total products delivered.

We also assume that the distance between locations will be taken as the Euclidean distance, and will as such be symmetric and fulfill the triangle inequality $c_{ij} \leq c_{ik} + c_{kj}$ for all $i, j, k \in N$. This assumption means that we can equate "choosing a route which passes the customer" with "choosing to deliver to said customer". It also means we never have to pass a location more than once, since it will never be beneficial. We use this to define the set of possible arcs as

$$\mathcal{A} = \{(i, j), \text{ for } i \in \mathcal{N} \backslash \{n\}, \, j \in \mathcal{N} \backslash \{0\} \text{ and } i \neq j\},$$

meaning a route can never return to the starting point, and never leave the end point once there.

To keep track of which arcs are used (and by extension which customers receive their products) we use binary variables

$$y_{ij} = \begin{cases} 1 & \text{if arc } (i, j) \text{ is used} \\ 0 & \text{else.} \end{cases}$$

We use variables x_i to keep track of the order in which we visit the customers, that is we demand that

$$x_j \geq x_i + 1, \text{ if } y_{ij} = 1, \ (i, j) \in \mathcal{A}. \tag{1}$$

We also introduce slack variables

$$\begin{aligned} s_K &= \text{unused capacity, and} \\ s_L &= \text{unused travel distance.} \end{aligned}$$

The model is given by

$$\begin{aligned}
\min \quad & s_K && && \text{(2a)} \\
s.t. \quad & \sum_{j \in \mathcal{N}} y_{ij} \leq 1 && \forall i \in \mathcal{N}, && \text{(2b)} \\
& \sum_{j \in \mathcal{N}} y_{ij} = \begin{cases} \sum_{j \in \mathcal{N}} y_{ji} & \forall i \in \mathcal{B}, \\ 1 & \forall i \in \{0, n\}, \end{cases} && && \text{(2c)} \\
& \sum_{i \in \mathcal{B}} d_i \sum_{j \in \mathcal{B}} y_{ij} + s_K = K, && && \text{(2d)} \\
& \sum_{i,j \in \mathcal{A}} c_{ij} y_{ij} + s_L = L, && && \text{(2e)} \\
& x_i - x_j + 1 \leq (n-1)(1 - y_{ij}), && \forall (i,j) \in \mathcal{A}, && \text{(2f)} \\
& x \geq 0 \in \mathbb{R}^n, \quad s_K \geq 0, \quad s_L \geq 0, && && \text{(2g)} \\
& y_{ij} \in \{0,1\}^{\mathcal{A}}. && && \text{(2h)}
\end{aligned}$$

a) Equivalent to maximizing the number of products delivered, the objective function seeks to minimize the unused capacity;

b) constraint (2b) ensures we only visit each customer once;

c) (2c) states that if you travel *to* a customer, you must also *leave* that customer. The start and end points are handles separately;

d) (2d) sets the value of s_K, and together with $s_K \geq 0$ ensures that the route does not exceed maximum capacity. We have chosen to consider a customer visited if we have traveled *to* them, rather than *away* from them, but the choice is arbitrary;

e) similarly (2e) sets the values of s_L, and together with $s_L \geq 0$ ensures that the route does not exceed maximum distance traveled;

f) constraint (2f) correspond to Eq. (1). Important to note is that while constraint (2f) keeps track of the order in which we visit the nodes, it - more importantly - ensures that the optimal route cannot contain contain any cycles. This together with (2c) ensures that the final route is connected.

Comments

The reason for including slack variable s_L is because it offers an easy way of taking the distribution cost into consideration. While weighting distribution cost against profit from delivery lies outside the scope of this project, expanding the objective to include the term $-\varepsilon s_L$ for a positive $\varepsilon <<$, means (with some caveats regarding the size of ε) that for the nodes included in the optimal route we choose the most efficient route.

Pre-processing of data, such as removing locations where $d_i > K$, etc. is recommended but has been left out of the discussion.