Warehouse Location Optimization

Ting-Hung Tsai

1 Aggregation

To simplify the calculation, one approach is to scale down the number of candidate warehouses and customers. We first observe that the candidate locations and customer locations are the same in this case.

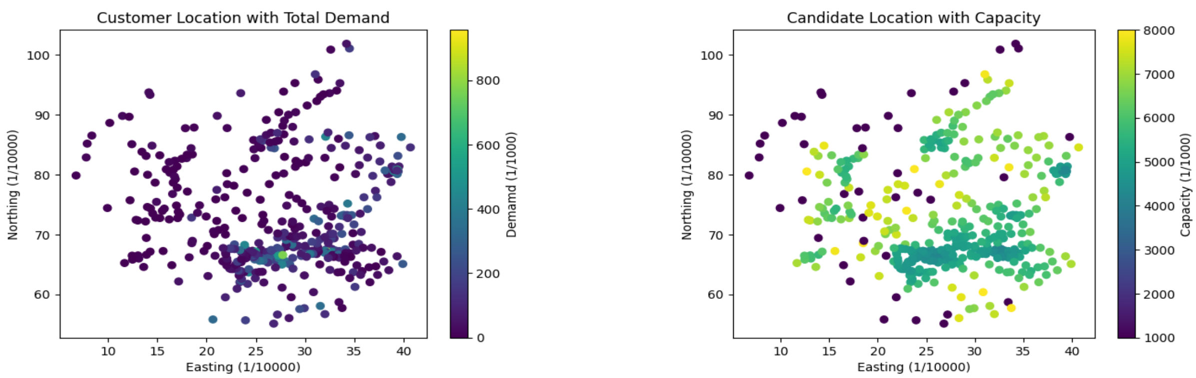

Figure 1: Customer locations with demand and candidate locations with capacity

As a result, our approach involves using K-means clustering to group customer together and retain the candidate that is closest to customer group centroid. This ensures the candidate could serve the customers in that group with the lowest transportation cost. Then, we check if the retained candidate's capacity is sufficient enough to handle all the demand of the entire group of customers. If not, we move on adding the candidate which is second closest to the centroid to the retained set and repeat the process until we have a set of candidates which can serve the group's demand.

After analysing the silhouette scores, we decide to group the customers into 30 clusters. This allows us to strike a balance between reducing complexity while still retaining geographic information about the customer base. For each of these 30 clusters, we select the candidate that is closest to the cluster's centroid to represent the cluster's location, as illustrated in the Figure 2. Finally, we retain 33 candidates for further analysis since the initial pool of candidates is short to satisfy the demand. Therefore, we added three candidates that are the second closest to the respective centroid. Such new set of candidates would be used in the later two problems while implementing the model to Case Study.

In addition, we would aggregate the customer demand data for the last question to simplify the computation process and avoid the complexities of uncertainty. We combine customers located within 15 miles of each other by order of customer ID data. We start with AB10 as a representative and combine all customers within 15 miles of AB10, then repeat the process by following the order of customer ID data. It is important to note that we would not reconsider any customer that has already been combined with a previous one. For instance, while AB11 has already been combined with AB10, we would not set AB11 as a representative. By doing this, we eventually obtained 133 representative customers. Using this aggregated customer data would significantly reduce computation time but it is worth mentioning that the result may not be as reliable and precise as using the original customer dataset.

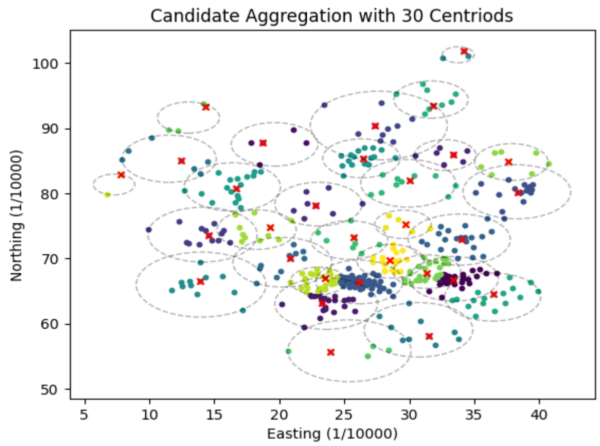

Figure 2: Aggregation of customer locations into 30 clusters

2 Deterministic Warehouse Location Problem

2.1 Mixed-integer Linear Programming

Data and Parameter

Let

i	be the index of customer; $i = 1, \ldots, I$ where I is the number of customers.
j	be the index of candidate warehouse; $j = 1, \ldots, J$ where J is the number of candidate warehouses.
k	be the index of supplier; $k = 1, \ldots, K$ where K is the number of suppliers.
t	be the index of year; $t = 1, \ldots, T$ where T is the number of years.
f_j	be setup cost of warehouse j.
g_j	be operating cost of warehouse j.
$\text{DemandPeriods}_{i,p,t}$	be the customer i's demand of product p on time t.
Capacity_j	be the maximal capacity of warehouse j.
$\text{SupplierCapacity}_k$	be the maximal capacity of supplier k.
$\text{SupplierProduct}_{k,p}$	be the binary indicator that whether supplier k is providing product p or not.
$\text{CostCandidateCustomer}_{j,i}$	be the cost of shipping every kilogram of products from warehouse j to customer i.
$\text{CostCandidateSupplier}_{j,k}$	be the cost of shipping every kilogram of products from supplier k to warehouse j.

Decision Variable

Let

$x_{i,j,p,t}$	be the percentage of customer i's demand of product p which is satisfied by warehouse j on time t.
$y_{j,t}$	be the binary indicator that whether warehouse j is operating on time t or not.
$z_{j,k,t}$	be the amount of product that supplier k is supplying to warehouse j on time t.

Constraints

Customers' demand has to be satisfied.

$$\sum_{j \in J} x_{i,j,p,t} = 1 \quad \forall i \in I, p \in P, t \in T$$

The quantity of demand delivered from a warehouse to all customers cannot be greater than its maximal capacity.

$$\sum_{i \in I} \sum_{p \in P} \text{DemandPeriods}_{i,p,t} x_{i,j,p,t} \leq \text{Capacity}_j y_{j,t} \quad \forall j \in J, t \in T$$

Once the warehouse is operating, it will keep running until the end.

$$y_{j,t} \geq y_{j,t-1}, \quad \forall j \in J, t \in T \backslash \{1\}.$$

The quantity of product delivered from a supplier to all warehouses cannot be greater than its maximal capacity.

$$\sum_{j \in J} z_{j,k,t} \leq \text{SupplierCapacity}_k \quad \forall k \in K, t \in T$$

The total quantity of product delivered from suppliers to a warehouses cannot be less than customers' demand that assigned to the warehouse.

$$\sum_{k \in K} \text{SupplierProduct}_{k,p} z_{j,k,t} \geq \sum_{i \in I} \text{DemandPeriods}_{i,p,t} x_{i,j,p,t} \quad \forall j \in J, p \in P, t \in T$$

The quantity of product delivered from all supplier to a warehouse cannot be greater than maximal capacity of that warehouse.

$$\sum_{k \in K} \text{SupplierProduct}_{k,p} z_{j,k,t} \leq \text{Capacity}_j y_{j,t} \quad \forall j \in J, p \in P, t \in T$$

The decision variables obey non-negativity or binary limit.

$$\begin{aligned} x_{i,j,p,t} &\geq 0 \\ y_{j,t} &\in \{0,1\} \\ z_{j,k,t} &\geq 0 \end{aligned}$$

Objective Function

There are three main costs in this problem: setup cost, operating cost and transportation cost.

The setup cost of warehouses is

$$\sum_{j \in J} f_j y_{j,T}.$$

Note that the index of y is T, i.e., specifically the last year. The reason is that the warehouse would continue running until the end so as long as we know which warehouse is operating in the last year, it must require a set up cost.

Operating cost of warehouses is

$$\sum_{j \in J} \sum_{t \in T} g_j y_{j,t}.$$

There are two parts of transportation costs, one is the delivery service cost between warehouses and customers,

$$\sum_{i \in I} \sum_{j \in J} \sum_{p \in P} \sum_{t \in T} x_{i,j,p,t} \text{DemandPeriods}_{i,p,t} \text{CostCandidateCustomer}_{j,i},$$

and the other is transportation cost between suppliers and warehouses,

$$\sum_{j \in J} \sum_{k \in K} \sum_{t \in T} z_{j,k,t} \text{CostCandidateSupplier}_{j,k}.$$

We simply add these two up as a transportation cost.

Thus, the objective function is the sum of above costs.

$$\text{Total Cost} = \text{Setup Cost} + \text{Operating Cost} + \text{Transportation Cost}$$

We aim to minimise the total cost.

$$\begin{aligned} \min & \sum_{j \in J} f_j y_{j,T} + \\ & \sum_{j \in J} \sum_{t \in T} g_j y_{j,t} + \\ & \sum_{i \in I} \sum_{j \in J} \sum_{p \in P} \sum_{t \in T} x_{i,j,p,t} \text{DemandPeriod}_{i,p,t} \text{CostCandidateCustomer}_{j,i} + \\ & \sum_{j \in J} \sum_{k \in K} \sum_{t \in T} z_{j,k,t} \text{CostCandidateSupplier}_{j,k} \end{aligned}$$

2.2 Results

Having applied the model to the Case Study, we now discuss the numerical results obtained from the model. The optimal solution to the problem is to set up warehouse FK21 in the beginning, followed by warehouses AB53 and DG11 in the second year, warehouse PH9 in year 4, warehouse TD3 in year 5, warehouse PA27 in year 7 and warehouse PH31 in year 8 with no new warehouses set up in the last two years. The total cost is £43130428.17 encompassing setup cost of £9086000, the operating cost of £6163600 and the transportation cost of £27880828.17.

Year	1	2	3	4	5	6	7	8	9	10
Warehouse	FK21	AB53, DG11	-	PH9	TD3	-	PA27	PH31	-	-

Table 1: New warehouses set up schedule in the optimal solution case

Obtaining the optimal solution takes up to 2.5 hours so we also explore suboptimal solutions that take less time in Table 2.

	Case 1	Case 2	Case 3	Case 4	Optimal Case
Approximate Run time (second)	600	900	1800	3600	9000
Year 1	EH30	EH30	FK21	FK21	FK21
Year 2	AB53, PA13	FK21	AB53	AB53, DG11	AB53, DG11
Year 3	FK21	AB53, PA13	-	-	-
Year 4	TD3	-	DG11	PH9	PH9
Year 5	-	DG11, PH9	PH9	TD3	TD3
Year 6	DG11, PA27, PH9	-	-	-	-
Year 7	-	-	PA27, PH31	PA27	PA27
Year 8	-	PA27	-	-	PH31
Year 9	-	-	PA38	-	-
Year 10	PA38	-	-	PA38	-
Setup Cost (£)	17694000	14476000	7252000	9058000	9086000
Operating Cost (£)	13728800	11426000	4534600	6016800	6163600
Transportation Cost (£)	20930312.85	22474268.19	32197353.27	28109032.4	27880828.17
Total Cost (£)	52353112.85	48376268.19	43983953.27	43183832.4	43130428.17
MIP Gap (%)	20.43	13.17	4.18	1.99	0

Table 2: Comparison between solutions that take different time

3 Stochastic Warehouse Location Problem

3.1 Mixed-integer Linear Programming

Data

Let

i	be the index of customer; $i = 1, \dots, I$ where I is the number of customers.
j	be the index of candidate warehouse; $j = 1, \dots, J$ where J is the number of candidate warehouses.
k	be the index of supplier; $k = 1, \dots, K$ where K is the number of suppliers.
t	be the index of year; $t = 1, \dots, T$ where T is the number of years.
s	be the index of scenarios; $s = 1, \dots, S$ where S is the number of scenarios.
f_j	be setup cost of warehouse j.
g_j	be operating cost of warehouse j.
Capacity_j	be the maximal capacity of warehouse j.
$\text{SupplierCapacity}_k$	be the maximal capacity of supplier k.
$\text{SupplierProduct}_{k,p}$	be the binary indicator that whether supplier k is providing product p or not.
$\text{CostCandidateCustomer}_{j,i}$	be the cost of shipping every kilogram of products from warehouse j to customer i.
$\text{CostCandidateSupplier}_{j,k}$	be the cost of shipping every kilogram of products from supplier k to warehouse j.
$\text{DemandPeriodScenarios}_{i,p,t,s}$	be the customer i's demand of product p on time t in scenario s.

Decision Variable

Let

$x_{i,j,p,t,s}$	be the percentage of customer i's demand of product p which is satisfied by warehouse j on time t in scenarios s.
$y_{j,t}$	be the binary indicator that whether warehouse j is operating on time t or not.
$z_{j,k,t,s}$	be the amount of product that supplier k is supplying to warehouse j on time t in scenarios s.

Constraints

Similarly to Task 1, we still have the constraint that once the warehouse is operating, it will keep running until the end.

$$y_{j,t} \geq y_{j,t-1}, \quad \forall j \in J, t \in T \backslash \{1\}.$$

For each warehouse in every periods and potential scenarios, the quantity of total products delivered to corresponding serviced customers cannot be greater than warehouse's maximal capacity.

$$\sum_{i \in I} \sum_{p \in P} x_{i,j,p,t,s} \text{DemandPeriodScenarios}_{i,p,t,s} \leq \text{Capacity}_j y_{j,t} \quad \forall j \in J, t \in T, s \in S$$

For each supplier in every periods and potential scenarios, the quantity of product delivered to corresponding supplied warehouses cannot be greater than supplier's maximal capacity.

$$\sum_{j \in J} z_{j,k,t,s} \leq \text{SupplierCapacity}_k \quad \forall k \in K, t \in T, s \in S$$

The quantity of product delivered from all supplier to a warehouse cannot be greater than maximal capacity of that warehouse.

$$\sum_{k\in K} \text{SupplierProduct}_{k,p} z_{j,k,t} \leq \text{Capacity}_j y_{j,t} \quad \forall j \in J, p \in P, t \in T$$

All customers' demand has to be satisfied in each period for every scenarios.

$$\sum_{j\in J} x_{i,j,p,t,s} = 1 \quad \forall i \in I, p \in P, t \in T, s \in S$$

The total quantity of product delivered from suppliers to a warehouses cannot be less than customers' demand that assigned to the warehouse.

$$\sum_{k\in K} \text{SupplierProduct}_{k,p} z_{j,k,t,s} \geq \sum_{i\in I} x_{i,j,p,t,s} \text{DemandPeriodScenarios}_{i,p,t,s} \quad \forall j \in J, p \in P, t \in T, s \in S$$

The decision variables obey non-negativity or binary limit.

$$\begin{aligned} x_{i,j,p,t} &\geq 0 \\ y_{j,t} &\in \{0,1\} \\ z_{j,k,t} &\geq 0 \end{aligned}$$

Objective Function

Similarly to Task 1, there are four main cost in this problems: setup cost, operating cost, service cost and transportation cost.

The setup cost of warehouse is

$$\sum_{j\in J} f_j y_{j,T}.$$

Operating cost of warehouse is

$$\sum_{j\in J}\sum_{t\in T} g_j y_{j,t}.$$

In each scenario, the delivery service cost between warehouses and customers is

$$\sum_{i\in I}\sum_{j\in J}\sum_{p\in P}\sum_{t\in T} x_{i,j,p,t,s} \text{DemandPeriodScenarios}_{i,p,t,s} \text{CostCandidateCustomer}_{j,i} \quad \forall s \in S$$

In each scenario, the transportation cost between supplier and warehouse is

$$\sum_{j\in J}\sum_{k\in K}\sum_{t\in T} z_{j,k,t,s} \text{CostCandidateSupplier}_{j,k} \quad \forall s \in S.$$

Thus, the objective function is the sum of above costs.

$$\text{Total Cost} = \text{Setup Cost} + \text{Operating Cost} + \mathbb{E}[\text{Delivery Service Cost}] + \mathbb{E}[\text{Transportation Cost}]$$

where

$$\mathbb{E}[\text{Delivery Service Cost}] = \frac{\sum_{s\in S} \text{Deliver Service Cost}_s}{S}$$

and

$$\mathbb{E}[\text{Transportation Cost}] = \frac{\sum_{s\in S} \text{Transportation Cost}_s}{S}$$

are the expected cost over all scenarios. We assume all scenarios are equally likely.

We aim to minimise the total cost.

$$\min \quad \sum_{j\in J} f_j y_{j,T} + \sum_{j\in J}\sum_{t\in T} g_j y_{j,t} + \frac{\sum_{s\in S} \text{Deliver Service Cost}_s}{S} + \frac{\sum_{s\in S} \text{Transportation Cost}_s}{S}$$

This model is quite time consuming, the number of variables that need to be optimized is multiples of scenarios. However, we can always get the greatest demands of every customers over all possible scenarios, which is

$$\text{DemandPeriods}_{i,p,t} = \max\{\text{DemandPeriodScenarios}_{i,p,t,s}\}_s = \text{DemandPeriods}_{i,p,t}(1+0.1t)$$

then use this data to run the model in Task 1 to get the most conservative results to save running time.

3.2 Results

The search for the optimal solution is constrained by the computational demands and the vastness of the data. Despite these challenges, the analysis reached a point of relative stability after 18,000 seconds of computation. This plateau indicates that additional computation is unlikely to result in significant alterations to the findings. Consequently, the results we present can be regarded as a reasonable approximation given the computational constraints.

The approximate optimal solution to the problem is to set up warehouse DD1 in the beginning, followed by FK21 in the second year. Warehouses AB14 and DG11 join in the third year, with TD3 in the fourth, PH9 in the fifth year, AB53 and PA27 in the sixth, PA38 in the seventh, PH31 in the eighth, IV28 and PH25 in the ninth year, and IV32 alongside KW5 in the last year. The total cost is £63671578.92, which includes £23786000 for setup, £14073400 for operations, and £25812178.92 for transportation.

Year	1	2	3	4	5	6	7	8	9	10
Warehouse	DD1	FK21	AB14, DG11	TD3	PH9	AB53, PA27	PA38	PH31	IV28, PH25	IV32,KW5

Table 3: New warehouses set up schedule in the approximate optimal solution case

	Case 1	Case 2	Case 3	Case 4	Case 5
Approximate Run time (s)	900	1800	3600	9000	18000
Year 1	AB53,DG11,FK21,HS8, IV22,IV28,IV44,IV55, KA10,KW5,PA27,PA38, PA45,PA68,PH9,PH25,PH31	DD1	DD1	DD1	DD1
Year 2	IV32	FK21	FK21	FK21	FK21
Year 3	-	AB14,DG11	AB14,DG11	AB14,DG11	AB14,DG11
Year 4	TD3	TD3	TD3	TD3	TD3
Year 5	-	PH9	PH9	PH9	PH9
Year 6	-	AB53,PA27	AB53,PA27	AB53,PA27	AB53,PA27
Year 7	-	PA38	PA38	PA38	PA38
Year 8	-	PH31	PH31	PH31	PH31
Year 9	-	IV28,PH25	IV28,PH25	IV28,PH25	IV28,PH25
Year 10	-	IV32,KW5	IV32,KW5	IV32,KW5	IV32,KW5
Setup Cost (£)	28641000	23786000	23786000	23786000	23786000
Operating Cost (£)	27714600	14073400	14073400	14073400	14073400
Transportation Cost (£)	22611352.88	25837791.94	25941303.07	25837308.54	25812178.92
Total Cost (£)	78966952.88	63697191.94	63706303.07	63696708.54	63671578.92
MIP Gap (%)	35.75	19.16	19.60	18.89	18.11

Table 4: Comparison between solutions that take different time

3.3 Less Conservative Model

The current model enforces the warehouses setup plan to satisfy all the possible scenarios, include some extremely high-demand scenarios which are very unlikely to happen. Therefore, we come up the following less conservative model and allow some customers' demand constraints to be violated. All the variables and parameter mentioned that are not defined below, stay the same with model in 3.1.

Decision variables

Let

$s_{i,j,p,t,s}$ be the slack variables indicate the percentage of customer i's demand of product p that delivered from warehouse j on time t in scenarios s.

Constraints

We allow some violation of customers' demand for every customer of each product on every time periods and scenarios. All the other constraints maintain the same.

$$\sum_{j\in J}(x_{i,j,p,t,s} + s_{i,j,p,t,s}) = 1 \quad \forall i \in I, p \in P, t \in T, s \in S$$

The slack variables obey non-negativity.

$$s_{i,j,p,t} \geq 0$$

Objective function

The objective function we want to minimise is similar of previous model and combine with the sum of slack variables times the demand which is the total violation of customers' demands.

$$\begin{aligned}
\min & \sum_{j\in J} f_j y_{j,T} + \\
& \sum_{j\in J}\sum_{t\in T} g_j y_{j,t} + \\
& \frac{\sum_{s\in S} \text{Deliver Service Cost}_s}{S} + \\
& \frac{\sum_{s\in S} \text{Transportation Cost}_s}{S} + \\
& \mu \sum_{i\in I}\sum_{j\in J}\sum_{p\in P}\sum_{t\in T}\sum_{s\in T} s_{i,j,p,t,s}\text{DemandPeriodScenarios}_{i,p,t,s}
\end{aligned}$$

where the parameter μ can be tuned to adjust the violation tolerance of customers' demand.

This is just an conceptual model without any code testing, and it is currently unknown what value of μ will be most effective.