École des Ponts ParisTech

Operations research project

Authors :
Group 14
Senhao DUAN, Ping'an YANG et Xiaohang HOU

8 juin 2022

1 Questions

1.1 The NP-hardness of the project question

To show that the air liquid question is NP-hard, we take advantage of the fact that facility location problem with constant construction cost is NP-hard. Facility location problem is formulated as following :
Input : A finite site $\mathcal{D}$ of customers, a finite site of potential facilities $\mathcal{F}$, a fixed cost $f_i \in \mathbb{R}_+$ for opening each facility $i \in \mathcal{F}$, and a service cost $c_{ij} \in R_+$ for each $i \in \mathcal{F}$ and $j \in \mathcal{D}$.
Output : Find a subset X of facilities and an assignment $\sigma : \mathcal{D} \to X$ of the customers to open facilities, such that the sum of facility costs and service costs

$$\sum_{i \in X} f_i + \sum_{j \in \mathcal{D}} c_{\sigma(j)j}$$

is minimum. We are going to write this as a air liquid problem. To do this, for every $(\mathcal{F}, \mathcal{D}, f_i, c_{ij})$ given in facility problem, we set $S = \mathcal{F}$, $I = \mathcal{D}$ and $d_i = 1$. Then we put different costs as follows.
Center building costs

$$building\_cost(s,x) = \begin{cases} f_i + a_s c_A^b & \text{if } s \in P \\ c_D^b & \text{if } s \in D, \end{cases}$$

where $c_A^b = c_D^b = +\infty$.
Production costs

$$production\_cost(i,x) \equiv 0.$$

Routing costs

$$routing\_cost(i,x) = \begin{cases} d_i c_2^r \Delta(s_i, i) & \text{if } s_i \in P \\ d_i[c_1^r \Delta(P_{s_i}, s_i) + c_2^r \Delta(s_i, i)] & \text{if } s_i \in D, \end{cases}$$

where $\Delta(s_i, i) = \frac{c_{ij}}{c_2^r}$, $\Delta(P_{s_i}, s_i) = 1$ and $c_1^r = +\infty$
Capacity costs
For this kind of costs we can simply assume the capacity $u_P = u_A = +\infty$ and hence we have

$$capacity\_cost(i,x) \equiv 0.$$

Therefore any facility location problems can be transformed to a air liquid problem in a polynomial time, which is equivalent to say that air liquid problem is NP-hard.

1.2 Reformulate as a MILP

We reformulate the air liquid problem as the following MILP :
Variables :
$upperbounds = 10^{11}$
$a\_s \in \{0,1\}$: the decision variables for the auto-production center.
$is\_pc[i] \in \{0,1\}$, where $i \in S$: the decision variables for the production center.
$is\_dc[i] \in \{0,1\}$, where $i \in S$: the decision variables for the distribution center.
$is\_connect[i][j] \in \{0,1\}$, where $i \in S$, $j \in S$: the decision variables for the connection of production center i and distribution center j.

$is\_connected\_pcclient[i][j] \in \{0,1\}$, where $i \in S$, $j \in I$: the decision variables for the connection of production center i and client j.
$is\_connected\_dcclient[i][j] \in \{0,1\}$, where $i \in S$, $j \in I$: the decision variables for the connection of distribution center i and client j.
$producing\_client[i][j] \in \{0,1\}$, where $i \in S$, $j \in I$: the decision variables for the connection of auto-production center i and client j.
$producing\_cost\_d[i][j] \in \{0,1\}$, where $i \in S$, $j \in I$: the decision variables for the indirect connection of auto-production center i and client j.
$is\_over\_cap[i] \in 0,1$, where $i \in S$: the variable that indicates whether the production center i is over capacity.
$\sigma[i][j] \in \mathbb{R}_+$ where $i \in S$ $j \in I$ the quantities of production which is exceeded the capacity.
$cost\_indirect[i][j] \in \mathbb{R}_+$ where $i \in S$ $j \in I$ the indirect production (demand from distribution center).
$cost\_cap[i]$, where $i \in S$.
$cost\_pre[i]$, where $i \in S$.
Constrains :
$\forall i,\ is\_pc[i] + is\_dc[i] = 1$,
$\forall j,\ \sum_i is\_connect[i][j] - is\_dc[j] = 0$,
$\forall i,\ \sum_j is\_connect[i][j] - is\_pc[i] \leq 0$,
$\forall i,\ j\ is\_connected\_pcclient[i][j] - is\_pc[i] \leq 0$,
$\forall i,\ j\ is\_connected\_dcclient[i][j] - is\_dc[i] \leq 0$,
$\forall j\ \sum_i is\_connected\_pcclient[i][j] + is\_connected\_dcclient[i][j] = 1$
$\forall i,\ j$:

$producing\_client[i][j] - upperbound \times a\_s[i] \leq 0$
$producing\_client[i][j] - is\_connect\_pcclient[i][j] \leq 0$
$producing\_client[i][j] + upperbound - upperbound \times a\_s[i] - is\_connect\_pcclient[i][j] \geq 0$

$\forall i,\ j,\ k$:

$producing\_cost\_d[i][j] + is\_connect\_pcclient[i][j] \leq 1$
$producing\_cost\_d[i][j] + 2 - is\_connect\_dcclient[k][j] - is\_connect\_pcclient[i][k] - a\_s[i] \geq 0$
$producing\_cost\_d[i][j] - is\_connect\_dcclient[k][j] \leq 0$
$producing\_cost\_d[i][j] - is\_connect\_pcclient[i][k] \leq 0$
$producing\_cost\_d[i][j] - a\_s[i] \leq 0$

$\forall j,\ k$

$\sigma[j][k] - is\_pc[j] \leq 0$
$\sigma[j][k] + 1 - \sum_i is\_connect\_dcclient[i][k] - \sum_i is\_connect[j][i] \geq 0$

$\forall i,\ j$

$cost\_indirect[i][j] \geq 0$
$cost\_indirect[i][j] - client\_demand[j] \leq 0$
$cost\_indirect[i][j] - upperbound \times \sigma[i][j] \leq 0$
$cost\_indirect[i][j] - client\_demand[j] - upperbound \times \sigma[i][j] + upperbound \geq 0$

$\forall i$

$\sum_j is\_connect\_pcclient[i][j] \times client\_demand[j] + cost\_indirect[i] - capacity\_normal - capacity\_bonus \times a\_s[i] - upperbound \times is\_over\_cap[i] = 0$

$\forall i$

$cost\_pre[i] - is\_over\_cap[i] \leq 0$
$cost\_pre[i] - a\_s[i] \leq 0$
$cost\_pre[i] - a\_s[i] + upperbound - upperbound \times is\_over\_cap[i] \geq 0$

$\forall i$

$cost\_cap[i] - upperbound \times is_over_cap[i] \leq 0$

$cost\_cap[i] - capacity\_costs \times \sum_j (is\_connect\_pcclient[i][j] \times client\_demand[j]) + cost\_indirect[i] - capacity\_normal \times is\_over\_cap[i] - capacity\_bonus \times cost\_pre[i]) = 0$

Objective function

$$
\begin{aligned}
COST = \sum_i & (BUILDINGCOST\_PLANT \times is\_pc[i] \\
& + BUILDINGCOST\_DISTRIBUTION \times is\_dc[i] \\
& + BUILDINGCOST\_AUTO \times a\_s[i]) \\
& + \sum_{i,\ j} (client\_demand[j] \times is_connect\_pcclient[i][j] \times PRODUCTION\_COST \quad + client\_d \\
& + client_demand[j] \times is\_connect\_dcclient[i][j] \times DISTRIBUTION\_COST \\
& - client\_demand[j] \times PRODUCTION\_AUTO \times producing\_cost[i][j] \\
& - client\_demand[j] \times PRODUCTION\_AUTO \times producing\_cost\_d[i][j]) \\
& + \sum_{i,\ j} (client\_demand[j] \times is\_connect\_pcclient[i][j] \times sites\_clients\_distance[i][j] \\
& \times ROUTE\_SECONDARY + client\_demand[j] \times is\_connect\_dcclient[i][j] \\
& \times sites\_clients\_distance[i][j] \times ROUTE\_SECONDARY) \\
& + \sum_{i,\ j,\ k} (client\_demand[j] \times is\_connect[i][k] \times sites\_sites\_distance[i][k] \\
& \times ROUTE\_PRIMARY) \\
& + \sum_i cost\_cap[i]
\end{aligned}
$$

1.3 Solve the instance KIRO-small

To solve the "KIRO-small" problem, we called the "PuLP module" in Python to build the MILP model and called "CPLEX solver" to solve this problem. And we obtained the following results.

Results for KIRO-small	
Integer optimal solution:	Objective = 3.2629393773e+07
Dual simplex - Optimal:	Objective = 3.2629393773e+07
Optimality:	100%

Figure 1 – Results for KIRO-small

2 Algorithm

2.1 Design of the algorithm

Through a preliminary observation of the original data of the "KIRO-large" problem, we find that there are 60 sites and 1470 clients involved. The scale is so large that the computational power of the computer is not sufficient to solve it directly by branch-and-bound method. Since this problem is a facility location problem, the main difference between the variables is the difference in their geographical locations, i.e., the difference in their latitudes and longitudes. Therefore, once the geographical distance between a "client A" and a "production center B" is large, it is impossible for the "client A" to be directly connected to the "production center B". And it is impossible for the "client A" to be indirectly connected to the "production center B" through another "distribution center C". Therefore, we consider dividing the original data by drawing a geographical partition line, in order to transform the original problem into smaller sub-problems. The geographic partition line only affects the results of some clients, production centers and distribution centers in a small area nearby, so the results obtained by this method should be acceptable. Based on this idea, we get the following pseudocode.

2.2 Pseudocode

Input : an instance of KIRO ;
Output : an optimal solution of minimum total cost x or $\varnothing$ if no optimal solution exist ;
Initialize : $L_1^1 \leftarrow \{(lat, lon)\}$ which contains the coordinates(latitudes and longitudes) of all clients, $x \leftarrow \varnothing, i \leftarrow 1$;
while $x = \varnothing$ **do**
  solve respectively the MILPs of the instance with branch-and-bound algorithm ;
  if all MILPs have optimal solutions **then**
    $x \leftarrow$ the optimal solutions ;
  else if i is odd **then**
    divide respectively $L_i^1, L_i^2, ..., L_i^{2^{(i-1)}}$ into two parts (with the average of the latitudes of clients as the dividing line) : $L_{i+1}^1, L_{i+1}^2, ..., L_{i+1}^{2^i}$;
    divide the previous MILPs into new sub-MILPs respectively including $L_{i+1}^1, L_{i+1}^2, ..., L_{i+1}^{2^i}$;
    $i \leftarrow i + 1$;
  else then
    divide respectively $L_i^1, L_i^2, ..., L_i^{2^{(i-1)}}$ into two parts (with the average of the longitudes of clients as the dividing line) : $L_{i+1}^1, L_{i+1}^2, ..., L_{i+1}^{2^i}$;
    divide the previous MILPs into new sub-MILPs respectively including $L_{i+1}^1, L_{i+1}^2, ..., L_{i+1}^{2^i}$;
    $i \leftarrow i + 1$;
  end if
end while
return : x

By calculation, we have $6n + n^2 + 5nm$ variables and $11n + 11nm + m + 5n^2m$ constrains, where $n = |S/2|$ and $m = |I/2|$ the upper bound of the complexity of our algorithm is $O(C_{n^2m}^{n^2+nm}(n^4m + n^3m^2))$

2.3 Numerical results

Results for KIRO-tiny	
Integer optimal solution:	Objective = 7.4309702000e+06
Dual simplex - Optimal:	Objective = 7.4309702000e+06
Optimality:	100%

Figure 2 – Results for KIRO-tiny

Results for KIRO-small	
Integer optimal solution:	Objective = 3.2629393773e+07
Dual simplex - Optimal:	Objective = 3.2629393773e+07
Optimality:	100%

Figure 3 – Results for KIRO-small

Results for KIRO-medium	
Integer optimal solution:	Objective = 6.9657843245e+07
Dual simplex - Optimal:	Objective = 6.9657843245e+07
Optimality:	100%

Figure 4 – Results for KIRO-medium

Results for KIRO-large	
Optimal solution:	Objective = 1.0282071139e+09
Bound:	1.0213959031e+09
Gap:	0.68112108e+07, 0.67%

Figure 5 – Results for KIRO-large