Package-Delivery

drop table group;
Junjie Wang 517021910093 dreamboy.gns@sjtu.edu.cn
Yikai Yan 517021910404 miku.miku.miku@sjtu.edu.cn
Wentao Qin 517021910483 423445630@qq.com

June 17, 2019

1 Abstract

We construct a graph to represent the city network. Problem 1 is NP. We solved it by modified $dijskra$. Problem 2 and 3 are also NP. We solved them by Sequential Algorithm. Problem 4$\leq_p$ Problem 1, but we solved it by approximation algorithm.

Contents

2 Symbol Table

We list the symbols we use to model the city network, commodities and orders in table 1 and 2. Detailed information is given later.

vertex	$city$	$index$, $attribute(small/substation/hub)$
edge	$dist$	
	$tools$	$depart\_city$, $arrival\_city$, $time\_on\_way$, $depart\_time$, $unit\_amount\_cost$

Table 1: City Network Model

commodity	$index$, $category$, $unit\_weight$
order	$seller\_city$, $purchaser\_city$, $order\_time$, $commodity\_index$, $commodity\_amount$, $emergency$

Table 2: commodity and order Model

3 Problem 1

3.1 Problem Analysis

In problem 1, since cities and transportation tools are not capacitied, orders are independent with each other. Therefore, we need to find an algorithm. The input is a particular order and the and the output is an optimized or approximated scheme for this order.

3.2 Modeling

- Modeling the city network.

  The city network is formulated into a graph. The vertices are the cities and the edges represent the transportation tools between 2 cities and the distance.

vertex u	edge (u, v)
$city$	$tools$
	$dist$

  $tools$ is the set of transportation tools from vertex u to v. $dist$ is the distance from vertex u to v (unit:km). In problem 1, a city is simply an index within set $[656]$ in our model.

- Modeling a transportation tool.

  Any transportation tool $tool \in tools$ is formulated into the set:

$$\{depart\_city\,,\, arrival\_city\,,\, time\_on\_way\,,\, depart\_time\,,\, unit\_amount\_cost\}$$

  $depart\_city$: the index of the city where the tool departs.

  $arrival\_city$: the city where it arrives.

  $time\_on\_way$: the time consumed on the way ($time\_on\_way = Average\_delay\_per\_trip + \frac{dist(u,v)}{speed}$,unit:$min$).

  $depart\_time$: the time it departs at.

  $unit\_amount\_cost$: the cost to transport a unit commodity from u to v ($unit\_amount\_cost = unit\_cost \times dist(u, v)$, unit:$\$/kg$).

- Modeling the commodity.

  A particular kind of commodity is formulated into the set:

$$\{index\,,\, category\,,\, unit\_weight\}$$

  $index$: commodity's index.

  $category$: the category which this kind of commodity belongs to.

  $unit\_weight$: the weight of one unit of such commodity (unit:kg).

  Notice that the information about commodity's unit price is omitted because this has nothing to do with our model.

- Modeling the order.

  Each order is formulated into the set:

$$\{seller\_city\,,\, purchaser\_city\,,\, order\_time\,,\, commodity\_index\,,\, commodity\_amount\,,$$
$$emergency\}$$

  $seller\_city$ and $purchaser\_city$: the index of the 2 cities.

  $order\_time$: the time when the order happens.

  $commodity\_index$: the index of the ordered commodity.

$commodity\_amount$: the number of commodities ordered.

$emergency$: set 1 if it's an emergency order and 0 else.

3.3 Problem Formulation

- Objective Function.

  The function is $f(p)$. p is a path from $seller\_city$ to $purchaser\_city$. We call this function overall evaluation.

  $$f(p) = a \times cost + b \times time.$$

  $cost$ is the total cost of the delivery scheme. And $time$ is the time that elapse from when the order happens to when the commodity arrives at the destination. Notice that customer's rating is proportional with $time$ and the lower the better. In our model we determine $a = 15$ and $b = 1$. I.e. f is the weighted average of $cost$ and $time$. $cost$ is with unit \$ and $time$ is with unit min.

  The constraint is that the scheme is represented by a simple path from $seller\_city$ to $purchaser\_city$. On the path, each 2 city is connected by exactly one transportation tool. $cost$ is calculated by simply add all the cost on each segment of the path.

  However, $time$ is hard to determine. Since the transportation tools has their $depart\_time$ and the commodity can't be transported immediately when it arrives. Therefore, $time$ can only be calculated by simulating the process. I.e. for each city u, if the commodity arrives earlier than the given transportation tool from u's $depart\_time$, add $time$ by $depart\_time - arrival\_time$. Else, it can only be transported out of city u next day. Hence, $time$ is added by $depart\_time - arrival\_time + 24 \times 60$.

- LP or ILP?

  We don't think this problem can be converted to LP or ILP. Since although the $cost$ function is linear, the $time$ function is not. As described above, $time$ function need to be calculated by simulation and it's discrete rather than continuous. Therefore, the objective function is not linear and we can't convert the problem to LP or ILP.

3.4 Complexity Analysis

This is a $NP\,Optimization$ problem $(l, sol, m, goal)$.

- l : $(G, orders)$. G is a given city network. $orders$ is a set of orders. This is poly-time recognizable.
- sol : $sol(G, orders)$ is a set of paths. For each $order \in orders$, there is a path from $seller\_city$ to $purchaser\_city$ in the set. Hence $|sol(x)| \leq p(|x|)$.
- m : $m(G, orders) = \sum_{order \in orders}(f(sol(order)))$. I.e. the sum of all the orders' schemes' overall evaluation. This function is poly-time computable.
- $goal$: minimize.

Therefore, this is a $NP\,Oprimation$ problem.

3.5 Algorithm Design

Although the time function is not linear and we can't find explicit weight function for edges, we can modify the classical $dijskra$ algorithm to generate a solution. The algorithm to solve one $order$ is shown in algorithm 1.

Algorithm 1: Schedule(G,$order$)

Input: city network G and an order $order$.
Output: The optimal path from $order.seller\_city$ to $order.purchaser\_city$

$OPT\_cities \leftarrow \{seller\_city\}$;
Assign each city's OPT_value to be ∞;
$seller\_city.arrival \leftarrow 0$;
$seller\_city.cost \leftarrow 0$;
$seller\_city.OPT\_value \leftarrow 0$;
$current\_city \leftarrow seller\_city$;
while $OPT\_cities \neq V(G)$ **do**
    foreach *city neighbor to* $current\_city$ **do**
        foreach $tool \in (current\_city, city).tools$ **do**
            $time\_tmp \leftarrow$ the reach time of $city$ via $tool$;
            $cost\_tmp \leftarrow current\_city.cost + tool.cost$;
            $value\_tmp \leftarrow a \times cost\_tmp + b \times time\_tmp$;
            if $value\_tmp < city.OPT\_value$ **then**
                $(city.arrival, city.cost, city.OPT\_value) \leftarrow (time\_tmp, cost\_tmp, value\_tmp)$;
                $city.pre \leftarrow current\_city$;
                $city.tool \leftarrow tool$;
    $current\_city \leftarrow$ the city in $cities/OPT\_cities$ with the lowest OPT_value;
    $OPT\_cities \leftarrow OPT\_cities \cup \{current\_city\}$;
$path \leftarrow$ the path recovered from pre and $tool$;
return $path$;

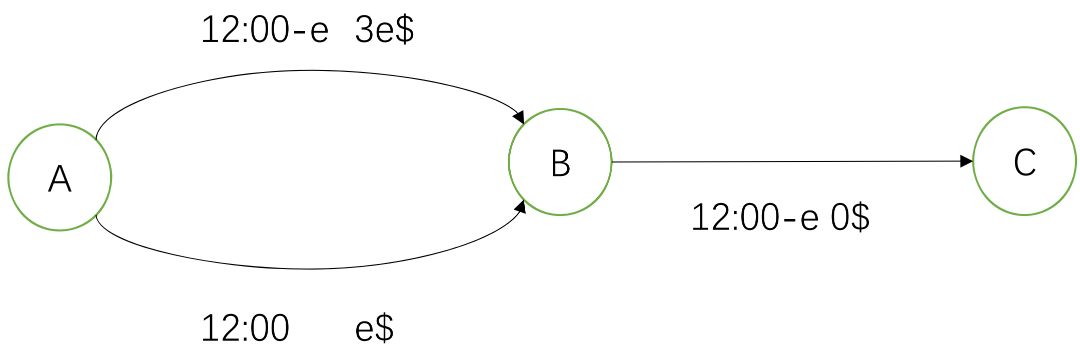

Figure 1: city network 1

3.6 Performance Analysis

1. Approximation Ratio

   Unfortunately, this algorithm doesn't have any approximation ratio theoretically. Consider the city network in figure 1. Suppose $a = b = 1$. $seller\_city = A$, $purchaser\_city = C$. The order happens at $12:00-e$. e's unit is min. The speed is infinite and delay is 0 ($time\_on\_way = 0$).

   $dijskra$ will choose the below edge to reach B from A, while the OPT chooses the above edge. $dijskra$ reaches C at $12:00-e$ the next day, using time $1440min$ and cost e$. OPT reaches C at $12:00-e$ this day, using time 0 and cost 3e$. Then we get

$$\frac{dijskra}{OPT} = \frac{1440+e}{0+3e} = \frac{1440+e}{3e}$$

   If we choose e arbitrarily small, the ratio is arbitrarily large.

2. Approximation Difference

   Then we use another criterion approximation difference to evaluate the lower bound of the performance. Approximation difference is defined as $\max\{f(dijskra) - f(OPT)\}$. This criterion is more appropriate for this problem than approximation ratio. This is because users care about how much time the package is late for. And the company cares about how much money they waste. They care the difference rather than the ratio.

   Using this criterion, the algorithm is $poly-apx$. I.e. $\max\{f(dijskra) - f(OPT)\}$ is a polynomial function of the input. Denote the number of cities as k. The approximation difference

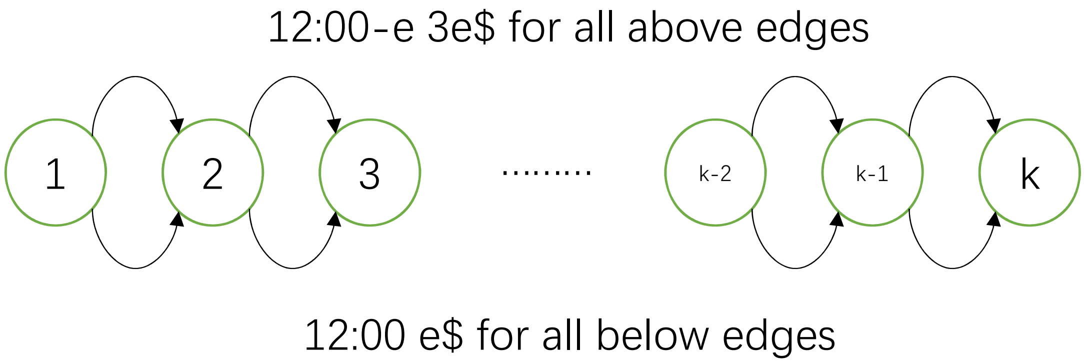

Figure 2: city network 2

is $1440b(k-1)$. Notice that the number of minutes in a day is 1440. Proof: if we don't care the waiting time, $dijskra$ is no worse than OPT. And the maximum time $dijskra$ waste on waiting is when it go through all k cities in a delivery and wait for about a day on each edge. Then the time wasted is $1440(k-1)$. This causes $1440b(k-1)$ difference to f.

This lower bound seems to be exaggerated. But it's actually a tight bound. See the network in figure 2. $time\_on\_way = 0$ for all edges and the order happens at $12:00-3$. OPT chooses the above edges for all, while $dijskra$ chooses the below ones. Then the $dijskra$ reaches k $k-1$ days later, while OPT reaches k immediately. If we choose e arbitrarily little, the approximation difference can be arbitrarily close to $1440(k-1)$.

3. Analysis in practice The $poly-apx$ seems to be terrible, but actually this algorithm works well in practice. This is because the real-world city network is a dense graph rather than the 1-dimension one shown in figure 2. And the probability that $dijskra$ chooses such a long path is nearly 0.

   In fact, we run the algorithm on the given city network. The largest f we get is

$$158885.46 = 18 \times 8676.70 + 1.2 \times 2254.05.$$

   We choose $a = 18$ and $b = 1.2$ in practice. This order costs 8760.70$ and uses about one day and a half. This is far lower than using $k-1$ days. Therefore, the algorithm works well in practice.

3.7 Efficiency Analysis

For each order, we run $dijskra$ once. Use $array$ for implementation. The algorithm checks each transportation tool at most once. Each check is $O(1)$. Also, each $edge$ and $vertex$ in G is checked at most once. Denote the total number of transportation tools in TableD-TransportationTools by $\#tools$. Denote the vertices set and edges set in G by V and E. Then the time complexity is $O(\#tools + |V| + |E|)$. Also, find the city with the lowest OPT_value is $O(|V|)$ in array implementation. Therefore, the time complexity is $O(\#tools + |V|^2)$. For each order, the algorithm runs once. Hence, the total time complexity is $O((\#tools + |V|^2) \times |orders|)$.

4 Problem 2

4.1 Problem Analysis

Assume the cost to construct a hub is $\hat{c}$.

$$\hat{c} = c_0 + c_1|orders|$$

c_0 is the cost for building the hub (a constant). $c_1|orders|$ is the cost for running the hub. $|orders|$ is the number of orders going through the hub. This is proportional with $|orders|$.

And assume the transportation cost between two hubs is $x(0 \leq x \leq 1)$ times that not between hubs. The f of a delivery is the sum of all the fs of its route.

For simplicity, **we assume that the delivery routes remain unchanged and only consider how to set up hubs and which transportation tool to use.** The idea is to sequentially check whether setting up a hub in a city can optimize $f(time, cost)$.

4.2 Problem Formulation and Complexity Analysis

We formulate this problem by $(I, sol, m, goal)$.

- I : $(G, orders, paths, \hat{c}, x)$. G is a city network. $orders$ is the set of orders. $paths$ is the scheme generated in problem 1. $paths$ is denoted by $|orders|$ paths in graph G. x and $\hat{c}$'s meaning are shown in section 4.1. This is poly-time recognizable.

- sol : $sol(G, orders, paths, \hat{c}, x)$ is the set of $(m\_paths, hubs)$. The requirements are:(i) the routes in m_paths is the same with that in $paths$ (transportation tools can be different); (ii) from one $hub \in hubs$ to one city, there is only one transportation tool used. This is also poly-time recognizable.

- m : The overall evaluation of scheme $paths$ is denoted by $sum\_f(paths)$. The overall evaluation of scheme m_paths is denoted by $sum\_f(m\_paths)$. Then $m(m\_paths, hubs) = sum\_f(paths) - sum\_f(m\_paths)$. I.e. the reduced amount of sum_f. This is poly-time calculatable.
- $goal$: maximize.

Therefore, we've shown that this is a NP-problem.

4.3 Algorithm Design

Since the orders and their routes have been solved in Problem 1, we have already known how many things of each type go through each city. We have also been restricted to using only one transportation tool from the hub to one city. So the transportation from the hub through different transportation tools has to be merged. Another check of which means of transportation is the best choice is therefore required. Algorithm 2 is the one checking whether construct a hub in city C. Algorithm 3 and 4 are 2 versions of the outer algorithms.

Algorithm 2: check(C, H, $paths$)

Input: City C, current hub set H, current scheme $paths$;
Output: The benefit from building a hub in city C, new scheme $paths$;

//Notice that when calculating sum_f, simulate the whole process, and take H into consideration;
$benefit \leftarrow 0$;
foreach *City C' reachable from C* **do**
  $max\_benefit \leftarrow 0$;
  foreach *$tool \in tools$ from C to C'* **do**
    $m\_paths \leftarrow paths$ but make all commodities from C to C' transported by $tool$;
    if $sum\_f(paths) - sum\_f(m\_paths) > max\_benefit$ **then**
      $max\_benefit \leftarrow sum\_f(paths) - sum\_f(m\_paths)$;
      $opt\_paths \leftarrow m\_paths$;
  $paths \leftarrow opt\_paths$;
  $benefit += max\_benefit$;
return $(benefit - \hat{c}, paths)$;

4.4 Performance Analysis

Notice that neither algorithm 2 nor 4 guarantees optimal solution. They are both sequential algorithms. Algorithm 2 sequentially determine the transportation tool from C to each city. Algorithm 4 sequentially determine whether constructing a hub in each city. However, the determination in previous iteration changes the scheme $paths$, thus influencing the later determinations. I.e. the order

Algorithm 3: $Set\_up\_hubs1(S, paths)$

Input: S a set of cities, scheme $paths$ got in problem 1
Output: H a set of hubs, new scheme $paths$;

$H \leftarrow \emptyset$;
for $c \in S$ **do**
  if $check(c, H, paths).benefit > 0$ **then**
    $H \leftarrow H \cup \{c\}$;
    $paths \leftarrow check(c, H, paths).paths$;
return *(H,paths)*;

Algorithm 4: $Set\_up\_hubs2(S, paths)$

Input: S a set of cities,scheme $paths$ in problem 1;
Output: H a set of hubs, new scheme $paths$;

$H \leftarrow \emptyset$;
while $H \neq S$ **do**
  $max\_benefit \leftarrow 0$;
  for $c \in S/H$ **do**
    if $check(c, paths).benefit > max\_benefit$ **then**
      $max\_benefit \leftarrow check(c, paths).benefit$;
      $max\_city \leftarrow c$;
  if $max\_benefit \leq 0$ **then**
    break;
  else
    $H \leftarrow H \cup \{max\_city\}$;
    $paths \leftarrow$ the scheme by setting max_city as a hub;
return *(H,paths)*;

in which the cities are checked actually influence the outcome. Hence, the optimal solution is not guaranteed.

We gave 2 versions of the outer algorithm. Algorithm 3 simply checks each city and determine whether to make it a hub. This may lead to an arbitrarily bad result. E.g. it chooses the first city with very little benefit. However this choice makes all other cities unable to be hubs ($benefit \leq 0$). But the OPT can achieve much more benefit. To make this situation impossible, we also design another algorithm 4.

Algorithm 4 is poly-apx ($k - apx$). k is the number of cities. The benefit achieved by algorithm 4 is at least the benefit of constructing the first max_city. I.e. $max\_benefit$ in the first iteration. The benefit achieved by OPT is at most $k \times max\_benefit$. I.e. constructing hubs in each city and each with benefit the same as $max\_benefit$. Therefore, we get the following formula:

$$\frac{OPT.benefit}{sequential.benefit} < \frac{k \times max\_benefit}{max\_benefit} = k.$$

Hence, algorithm 4 is poly-apx.

4.5 Efficiency Analysis

The function $check$ is $O(\#tools_C \times |E| \times |orders|)$. $\#tools_C$ is the number of transportation tools from city C. $|E|$ is the number of edges in graph G. This is because that the algorithm simulates to get benefit $\#tools_C$ times. And each simulation requires $O(|E| \times |orders|)$ time.

For algorithm 3, the time complexity is $O(\#tools \times |E| \times |orders|)$. $\#tools$ is the total number of transportation tools. This is because adding up $\#tools_C$ for each city reaches $\#tools$.

For algorithm 4, the time complexity is $O(\#tools \times |E| \times |orders| \times |V|)$. $|V|$ is the number of vertices in G. This is because we check each city for at most $|V|$ times.

Since algorithm 4 is much slower, we implemented algorithm 3 instead.

We can improve the efficiency by only checking the orders whose delivery scheme has changed. However, this won't change the complexity in the worst case. This is because in the worst case, all orders go through all cities. Hence, when the transportation tools are merged, all orders' delivery schemes changed. Anyway, this optimization works well in practice.

5 Problem 3

5.1 Problem Analysis

For this problem we assume that the hub and the substation can function in the same city since hubs are capacitied. This is quite similar to problem 2 except that a little change has to be made to the algorithm to check for each city. Obviously, this is still a NP problem.

5.2 Algorithm Design

Since the hubs are capacitied, we have to take this into consideration and decide which commodities are to be transported through hubs and which through substations.

First, we should determine which orders to be delivered by the hub in the city. To maximize the benefit from the hub, we can treat the time and money we save as the value and the order's commodity's weight as the weight. Then convert the problem into a knapsack problem. The algorithm 5 illustrates our idea.

Algorithm 5: $checkTool(C, C', T)$

Input: Cities C and C' in which C' is reachable from C by T, a transportation tool;
Output: The maximum benefit of delivery from C to C' if we build a hub in city C and use T to deliver the goods;

for *Commodity c going through C that are allowed in the hub of C and in T* **do**
  $value[c] = f(p') - f(p)$ in which p is the path with the old transportation while p' goes by the new one;
  $r[c] = \dfrac{value[c]}{weight[c]}$;
$b \leftarrow$ The capacity of the hub;
Sort r in non-increasing order $r_1, r_2, \ldots, r_k$;
$benefit \leftarrow 0$;
for $i \leftarrow 1$ *to* n **do**
  if $b \geq weight[c_i]$ **then**
    $c \leftarrow \max_c\{value_{max}, r_i\}$;
  else
    $c \leftarrow \max_c\{value\}$;
    if $b < weight[c_i]$ **then**
      $c \leftarrow null$;
  Put c into the hub;
  $benefit \leftarrow benefit + value[c]$;
  $b \leftarrow b - weight[c]$;
if $benefit$ < *the benefit of putting only the feasible order with the largest $value$* **then**
  Only put the feasible order with the largest $value$ instead;
return $benefit$;

Then we can give the modified algorithm $check$ in algorithm 6

Algorithm 6: check(C)

Input: City C;

Output: The overall benefits from building a hub in city C;

$benefit \leftarrow -\hat{c}$;

for *City* C' *reachable from* C **do**

 $c \leftarrow \min_{T \in \text{Available tools from } C \text{ to } C'} checkTool(C, C', T)$;

 $benefit \leftarrow benefit +$ (the current total f(time,cost) from$C$ to C') $- c$;

return $benefit$;

And the outer algorithm remains unchanged.

Notice that when calculating the benefit, let those glass-made or inflammable products transported via the substation rather than hub in those cities which don't allow them. Also, the original scheme needs to be changed. For those orders with liquid and inflammable products, run the algorithm in problem 1 again, adding the constraint. Then run algorithms on the modified original scheme. Anyway, these 2 constraints will not cause substantial difference to our model.

5.3 Performance Analysis

The approximation ratio changes to $2k$, which is still poly-apx. This ratio is got by multiplying the ratio in problem 2 by 2. In class, we are taught that the knapsack problem is with approximation ratio 2. Hence multiply the approximation ratio in the 2 procedure and we got the overall approximation ratio.

5.4 Efficiency Analysis

Use the symbols defined in section 4.5. The time complexity of algorithm 5 is $O((|E|+\log|orders|)|orders|)$. Calculating each order's value is $O(|E||orders|)$. Sorting the orders is $O(|orders|\log|orders|)$.

The complexity changes because each check takes more time. The check function's complexity becomes $O(\#tools_C(|E|+\log|orders|)|orders|)$.

Then the complete algorithm's time complexity is $O(\#tools(|E|+\log|orders|)|orders|)$ (algorithm 3) or $O(\#tools(|E|+\log|orders|)|orders||V|)$ (algorithm 4).

6 Problem 4

6.1 Problem Analysis

Here is our assumption for this problem. If the $seller\_city$ is not a substation, it should be first delivered to a $substation$. As soon as the commodity reaches any substation, it should be delivered between substations, i.e. not going back to small cities. Once the commodity goes back to some small city, it should be transported to $purchaser\_city$ without going back to any substations. I.e. the commodities should be transported first between small cities, then substations and then again small cities. This is close to the reality.

Moreover, in reality those substations should be reachable from each other without small cities as intermediate. Otherwise, some substations may become islets. And any small city should be reachable from and to at least one substation. Otherwise, the city becomes an islet. Therefore, we make these 2 as our requirements for the data set.

We can revise our model by only set those cities with substations as the vertices. For those small cities without substations, associate them with the nearest substation. I.e. construct "huge vertices".

6.2 Complexity Analysis

Actually $problem_4 \leq_p problem_1$. Since for the situation where $seller\_city$ and $purchaser\_city$ are neither substation. All feasible delivery delivers from $seller\_city$ to substation s and then to substation t and finally to $purchaser\_city$. (s can be the same substation as t). Then we can let the algorithm exhaust all the possibilities of the choices of s and t. There are at most $O(|cities|^2)$ possibilities. Therefore, running the algorithms this many time can solve problem 4.

6.3 Modeling

- Modeling cities

  A city is modeled by a pair

$$(index, attribute).$$

  $index$ is the city's index. $attribute$ can be $substation$ or $small$.

- Modeling graph

  In this problem, we construct 2 graphs G and G_r. G is the same graph with problem 1. For G_r, delete all the cities without substation from the G in problem 1. And delete corresponding edges.

6.4 Algorithm Design

Although exhausting all probabilities can reach an optimized solution, it's too inefficient. Therefore, we revise the algorithm to be algorithm 7. Notice that theoretically this may lead to arbitrarily bad result. E.g. $seller\_city$ is not a substation but $purchaser\_city$ is. However the nearest substation from $seller\_city$ is infinitely far from the $purchaser\_city$. However, this can never happen in reality since the triangle law on real-world map. Therefore, this algorithm can lead to very close answer to the OPT in practice.

Algorithm 7: Schedule(G,$order$)

Input: city network G and an order $order$.

Output: A path from $order.seller\_city$ to $order.purchaser\_city$

foreach $(u, v) \in E(G)$ **do**
  Weight the edge by $Weight(u, v, G, order).minweight$;
foreach $(u, v) \in E(G_r)$ **do**
  Weight the edge by $Weight(u, v, G_r, order).minweight$;
$G_{rr} \leftarrow$ graph G deleting all substations;
$substation\_s \leftarrow order.seller\_city$;
$substation\_t \leftarrow order.purchaser\_city$;
if $order.seller\_city.attribute == substation$ && $order.purchaser\_city.attribute == subtation$ **then**
  Run Algorithm 1 in problem 1 on G_r;
  return *the path got*;
if $order.seller\_city.attribute == small$ **then**
  Use $dijskra$ on G_{rr} to calculate the shortest path from $order.seller\_city$ to each small city;
  foreach *small city* **do**
    Update the shortest distance of neighbor substations;
  Choose the substation s_m with minimum weight of the shortest path;
  $substation\_s \leftarrow s_m$;
if $order.purchaser\_city.attribute == small$ **then**
  Use $dijskra$ on G_{rr} to calculate the shortest path from each small city to $order.purchaser\_city$;
  foreach *small city* **do**
    Update the shortest distance of neighbor substations;
  Choose the substation t_m with minimum weight of the shortest path;
  $substation\_t \leftarrow t_m$;
Run $dijskra$ algorithm G_r to find a shortest path from $substation\_s$ to $substation\_t$;
Combine the 3 paths as $path$;
return $path$;

The algorithm's complexity is also

$$T(|orders|, \#cities) = |orders| \times \#cities^2.$$

7 Performance Evaluation

7.1 Delivery Scheme

The entire result is available on http://resources.dbgns.com/package-delivery/results.

1. Problem 1

   We run the algorithm for every order and got their delivery schemes. We list some as follow.

   ```
   For order 0 with totalWeight: 187.529 and emergency: 0, Best strategy is:
   Paths: [655, 614, 593, 10], Vehicles: ['Truck', 'Truck', 'Truck'],
   AmountCost: 200.161$, TimeCost: 584.11m, ArrivalTime: 23:50:55

   For order 33 with totalWeight: 7.612 and emergency: 0, Best strategy is:
   Paths: [362, 621, 73, 646], Vehicles: ['Plane', 'Truck', 'Truck'],
   AmountCost: 18.661$, TimeCost: 685.48m, ArrivalTime: 23:33:39
   ```

2. Problem 2

   In this problem, the delivery scheme will change because of the constraints of hubs. And we only list the cities where we construct hubs here. 4 hubs are constructed in total.

   ```
   106 205 164 503
   ```

3. Problem 3

   Also, we list the cities where we construct hubs here. 32 hubs are constructed in total.

   ```
   358 400 290 406 549 417 529 577 187 106 51 308 651 531 183 252
   205 469 223 164 503 117 199 197 370 352 524 449 460 424 397 302
   ```

4. Problem 4

   Since the constraints added, the delivery schemes change. E.g. order 0's scheme changed to:

   ```
   For order 0 with totalWeight: 187.529 and emergency: 0, Best strategy is:
   Paths: [10, 548, 170, 54, 655, 403, 170],
   Vehicles: ['Truck', 'Truck', 'Ship', 'Truck', 'Truck', 'Truck'],
   AmountCost: 343.953$, TimeCost: 5398.69m, ArrivalTime: (+4day):8:05:30
   ```

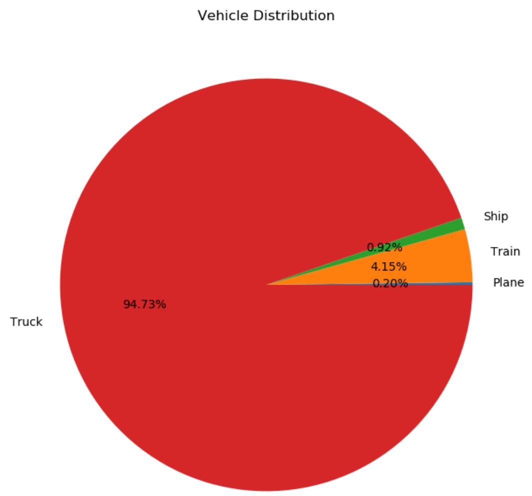

Figure 3: problem 1

7.2 Ratio of Transportation Tools

In this section, we visualized the ratio of transportation tools for each problem. See figure 3, 4, 5 and 6. From this figure, we found that truck occupied a huge ratio of chosen transportation tools. And plane occupies very little. One reason is that the number of trucks is far more than that of other transportation tools. Another reason is that plane is too expensive. And this also correspond with reality. Since in reality, only emergency packages will be transported by plane. Another interesting result is that, in problem 4, ship occupied a substantially larger ratio.

7.3 Trade-off Between Customer Rating and Cost

We changed parameters a and b. ($f = a \times cost + b \times time$.) Under different parameters, use a/b as independent variable. And use $cost$ and $time$ as dependent variables. We get the figure 7.

$AmountCost$ corresponds to $cost$ and $timeCost$ corresponds to $time$. From this figure, we found that as a/b increases, $time$ and $cost$ increases and decreases respectively. Therefore, there is actually a trade-off between customer rating and cost.

Since there is a trade-off between them, maybe we need to determine a and b more carefully to strike a balance between them. The objective is to achieve lower $time$ and $cost$. Hence, we use $time \times cost$ achieved by the algorithm as the dependent variable. And use a and b as independent variables. Then the relationship between them is shown in figure 8

In the figure weight of amount cost corresponds to a and weight of time cost corresponds to b.

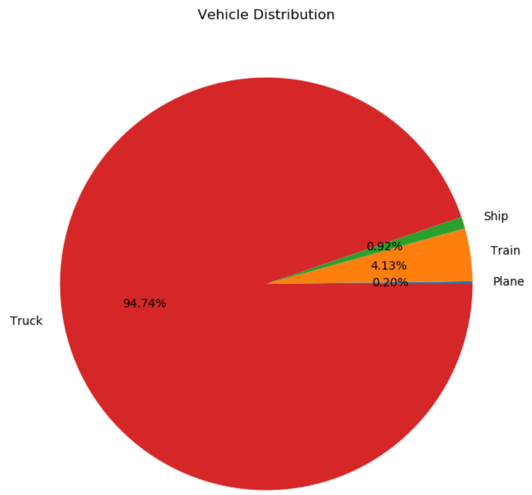

Figure 4: problem 2

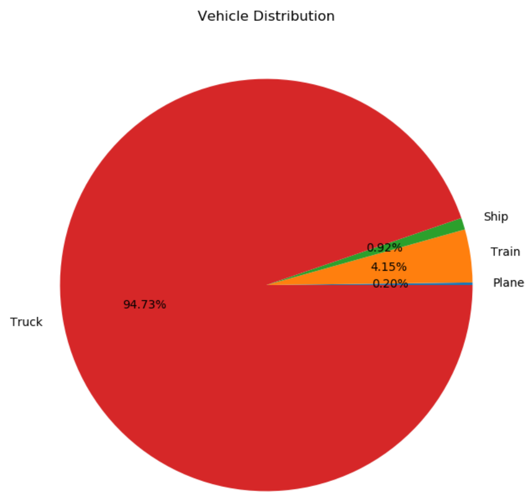

Figure 5: problem 3

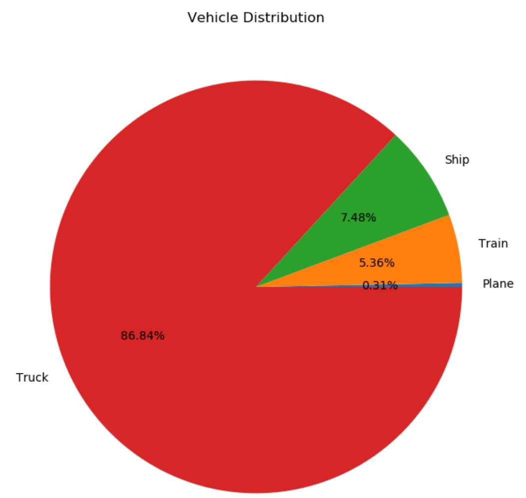

Figure 6: problem 4

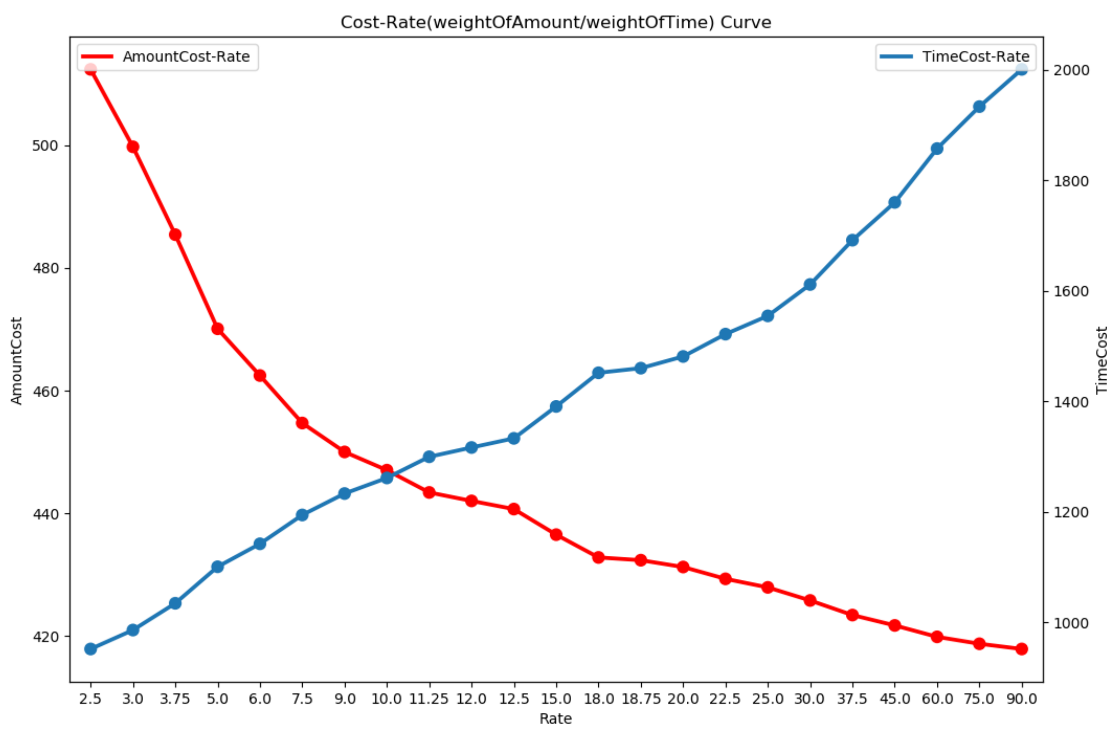

Figure 7: Trade of between customer rate and cost

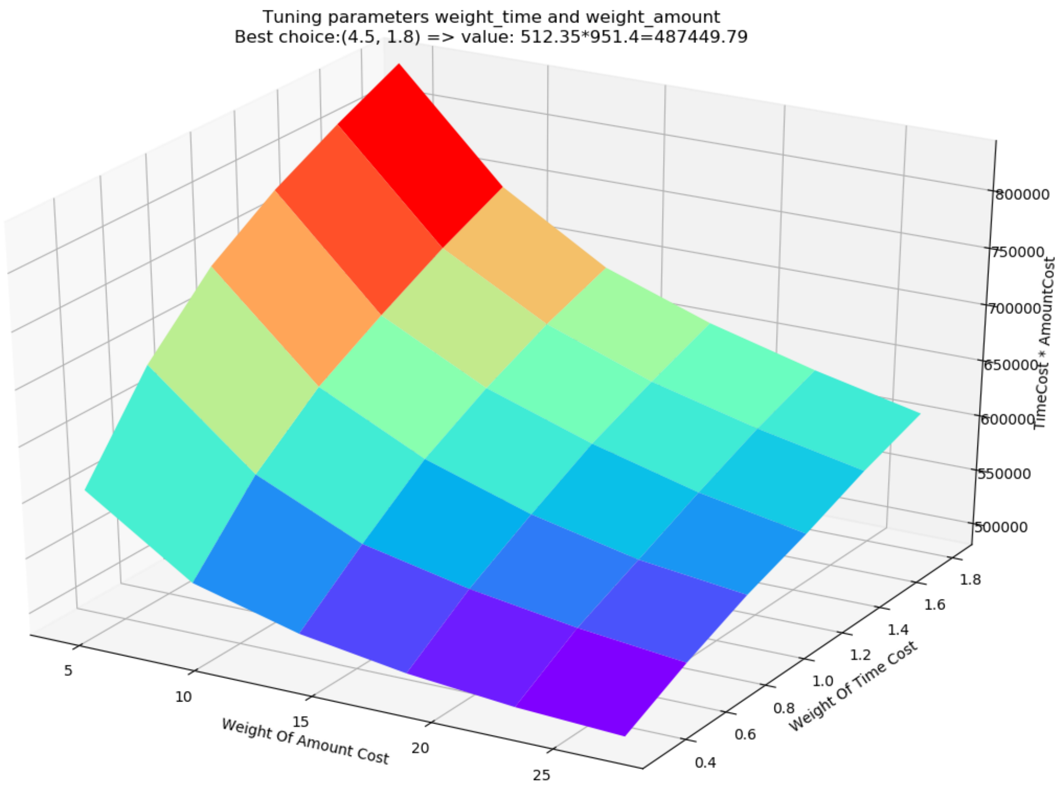

Figure 8: $time \times cost - (a, b)$

Then we can see from the figure that the best choice of a and b is 4.5 and 1.8 respectively.

7.4 Average Cost and Time

We list the average cost and time achieved in each problem here in table 3.

	average cost	average time	number of hubs
problem 1	439.50	1357.80	
problem 2	437.94	1369.31	4 hubs
problem 3	439.49	1359.79	32 hubs
problem 4	969.08	3374.58	

Table 3: Overall Evaluation

From the table, we found some results:

1. Problem 2 and 3 achieved lower $cost$ and $time$ than problem 1 because of the hubs.
2. In problem 3, much more hubs are constructed. This is because that this problem allows hub and substation exist in the same city.
3. In problem 4, average cost and time are much larger than in problem 1. This is because not all cities are substations. Hence, longer delivery route is required.

8 Sensitivity Test

In the program, we determine $c_0 = 3000$ and $c_1 = 300$. The 2 are parameters determining hub building cost. And we determine $x = 0.7$. This is the ratio for hub transportation. And $capacity = 1000$. This is the hub capacity.

In this section, we change the 4 parameters and observe their effect on our evaluation function $f = a \times cost + b \times time$.

1. c_0 and c_1

   The achieved f on different c_0 and c_1 is plotted in figure 9 and 10. Select 2 neighbor points around the value using in the program. And we find that the slopes are 0 for both figure. This is because small change to these 2 parameters will not influence how we construct the hubs. Therefore, the model is not sensitive to these 2 parameters. But of course, we can see also in the figures that large change to these 2 parameters will still influence the result.

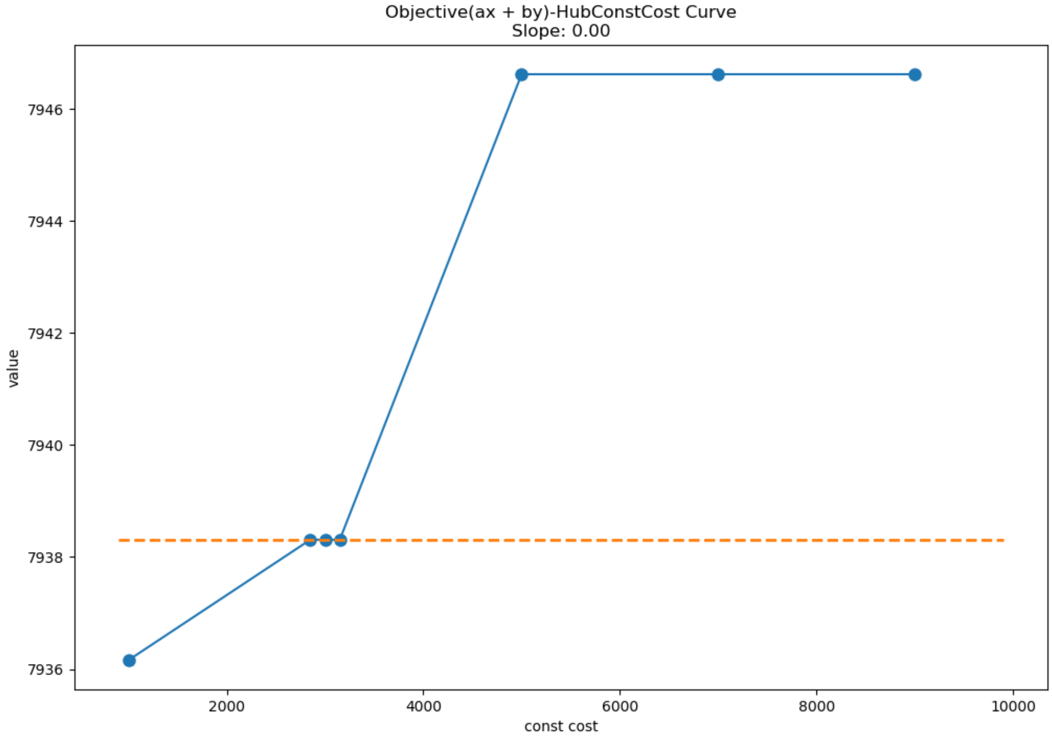
Objective(ax + by)-HubConstCost Curve
Slope: 0.00
value
7936
7938
7940
7942
7944
7946
const cost
2000
4000
6000
8000
10000

Figure 9: $f - c_0$

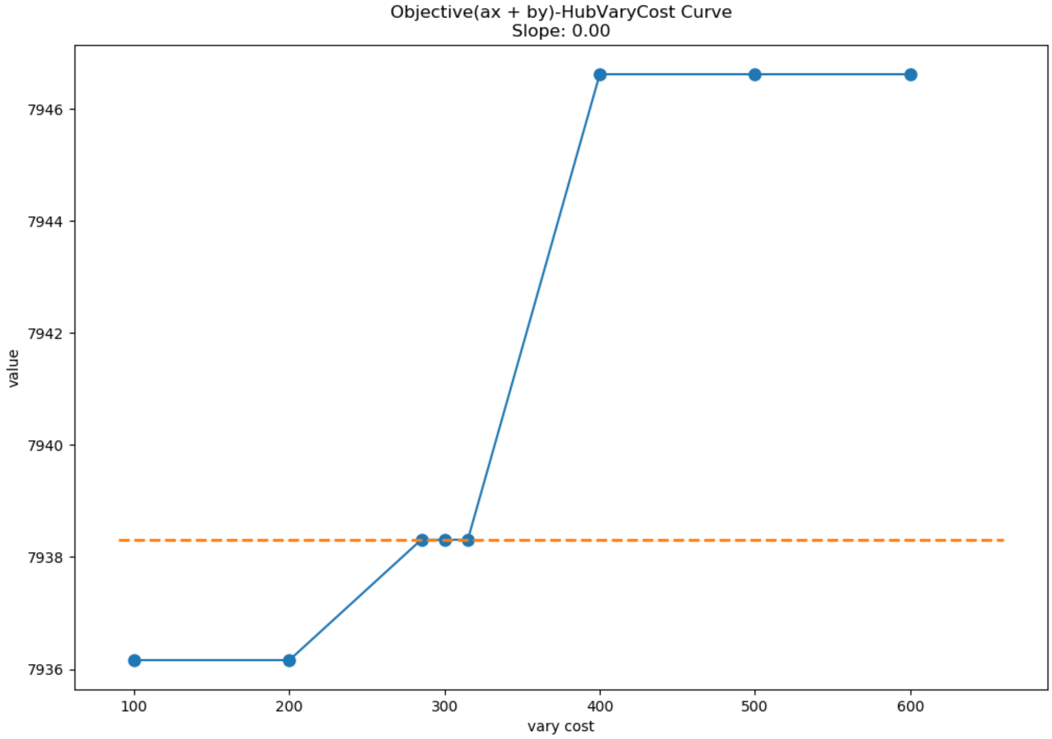

Figure 10: $f - c_1$

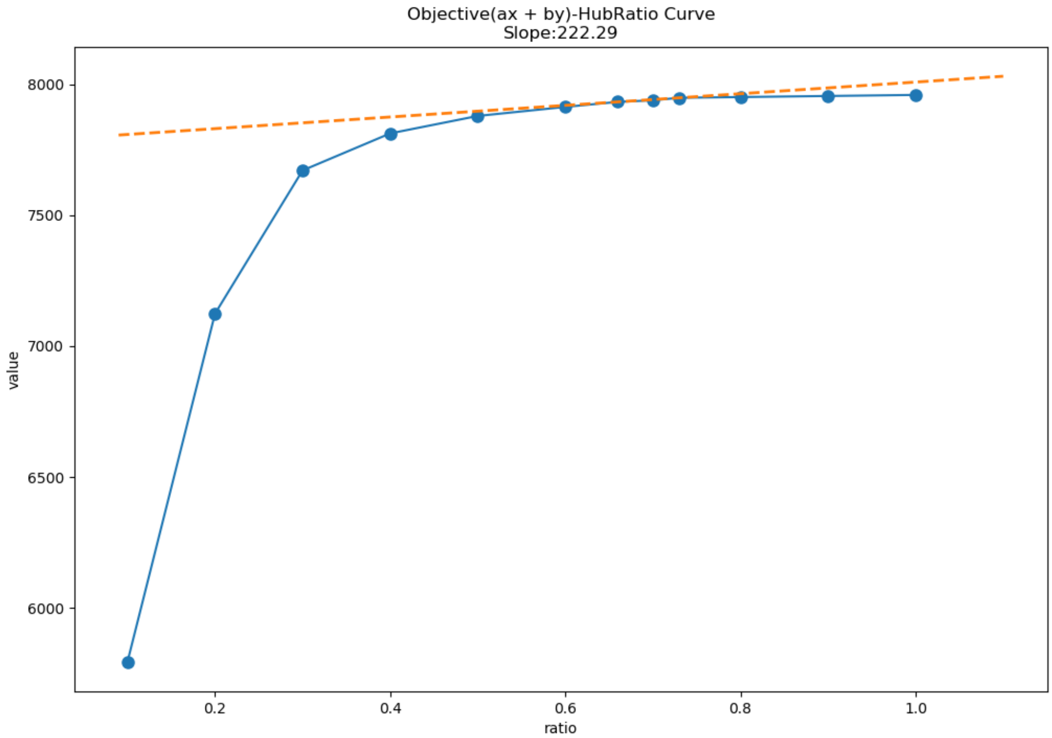

Figure 11: $f - x$

2. x The effect of x is shown in figure 11. We can see that the overall evaluation around $x = 0.7$ is relatively stable. I.e. our model is also not sensitive to x around our given value 0.7. However, around lower value of x, our model is much more sensitive.

3. $capacity$ In problem 3, we give each hub restricted capacity. Hence, we look into the effect of it also. The figure is given in figure 12. We can see that the slope is 0 around our given value 1000. This is also because small difference of parameters will not make changes to the hub construction. Hence, our model is not sensitive to $capacity$ as well.

9 Summary

We refer to shortest path problem, sequential algorithm and knapsack problem to solve this project. And we sum the advantages and disadvantages of our model as follow:

Disadvantages:

1. In the worst case, the algorithms lead to poly-apx result in problem 2 and 3. And theoretically the algorithm can lead to arbitrarily bad result in problem 4. This may be far from the optimal.

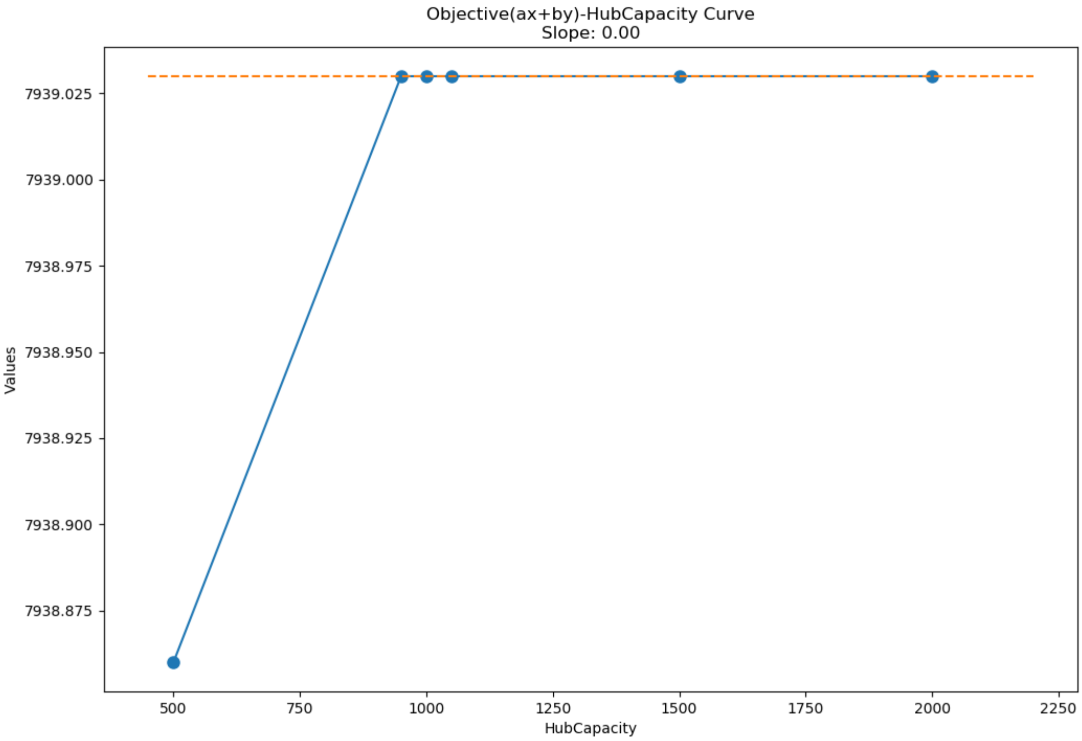

Figure 12: $f - capacity$

For problem 1, the "approximation difference" is also polynomial.

2. Although the algorithm is poly-time algorithm, it's still not efficient enough. Especially for problem 2 and 3, it simulates the whole process in the worst case for a check. This needs very much time.

Advantages:

1. The algorithm works out very close solution to OPT for problem 1 in practice. And it's efficient.
2. In practice, the algorithms performs well. Because of the triangle inequility of real-world maps, the algorithm for problem 4 works out schemes very close to the optimal ones. For problem 2 and 3, the worst case seldom happens in practice. Hence the efficiency and performance are much better.
3. All real-world objects (cities, transportation, $\cdots$) are all modeled to mathematical symbols. This helps the theoretical analysis and algorithm design.

10 Acknowledgement

From this project, we've learned very much. The largest achievement is that we've learned how to design algorithms in this project. And since we utilized the $dijskra$, sequential algorithm and $knapsack$ problem in this project, we become much more familiar to them. Also, as for the theoretical part, by proving the approximation ratio, we've got more familiar with the mathematical methods.

And from this project, we found that those problems in reality may seems simple. However, when we truly analyze them, investigate the algorithm, it turns out to be really hard. In problem 1, we find that the orders are independent with each other. And we thought we can solve the problem by processing the orders one by one. Then this becomes an easy problem. However, we struggled a lot to find such an algorithm and made many mistakes. And this problem even seems to be a NP complete one.

Thanks for professor Gao's instructions along the semester and the careful design of our labs and projects! We've really learned much from this course!

References

[1] Gao Xiaofeng. Slide12-ShortestPath. 2019

[2] Gao Xiaofeng. Slide15-NPReduction. 2019

[3] Gao Xiaofeng. Slide16-ApproximationI. 2019

[4] Gao Xiaofeng. Slide17-ApproximationII. 2019

Appendix

Code is available on github. Website: https://github.com/dbgns/package-delivery

Model is available on website: http://resources.dbgns.com/package-delivery/models

The detailed result is available on website: http://resources.dbgns.com/package-delivery/results