International Journal of Computer Applications (0975 - 8887)
Volume 67 - No. 11, April 2013

Approximation Algorithm for Facility Location Problems

Pawan Kumar Patel
Department of Computer Science & Engineering
Indian Institute of Technology, Kanpur, INDIA
Karnataka, India

ABSTRACT

Significant research effort has been devoted in the study of approximation algorithms for NP-hard problems. In this work we modify a known primal-dual approximation algorithm for facility location problem. Although we fail to give a performance guarantee for the new approach but we show that our method performs better in a tight case.

Keywords:

3D face recognition, range image, radon transform, Symbolic LDA.

1. INTRODUCTION

We will solve NP-hard problem in this work, namely the the *facility location* problem. Good approximation algorithms are known in the literature for this problem. Our objective is to seek possible improvements in one of the approximation algorithms for the facility location problem.

1.1 Facility location problem

An oil distribution company plans to set up a oil distributing station in some of the chosen locations (potential facility sites). The company has already figured out the cost of setting up the distribution station in each of these sites. The cost of supplying the oil from each site to each city is also known. The objective of this problem is to select some of the sites to setup the facilities so that the total cost of setting plus the cost of supplying to all the cities is minimized.
The above problem is an example of the facility location problem. Since the early 1960's location problems have important place in operation research. Facility location problems are helpful in taking decision to setup factories, hospitals, warehouse, fire station. Some of the significant contributions to this problem. include Stollsteimer [1], Kuehn and Hamburger [2], Manne [3].
A variant of the problem, called *metric uncapacitated facility location problem* is described as follows. L is a set of location where facilities may potentially be built. For every location information about the cost of building is given. C is a set of demand points(cities) each of which has to be assigned to a functional facility to receive service. A fixed one-time cost is known to connect a given city to a given facility location. The objective is to find a set of locations where facilities should be built and associate each city to one of the built facilities such that the total cost, the cost of facility building and the cost of connecting, is minimized.

Cornuejols, Nemahauser and Wolsey [4] showed that Uncapacited facility location problem *UFLP* is NP-hard. Their results extends to metric UFLP as well. Guha and Khullar [5] showed that metric UFLP is APX-complete. Sviridenko [6] proved that a performance guarantee of less than 1.467 can not be given for an approximation algorithm for metric UFLP unless $P = NP$.
The first approximation algorithm for this problem was due to Hochbaum [7] having $O(\log n)$ approximation guarantee. The first constant factor approximation was due to Shymoys et. al. [8] with 3.16 approximation guarantee. Subsequently Jain and Vazirani [9] gave a 3-factor approximation algorithm. Guha and Khullar [5] obtained 2.47 performance guarantee for this problem using Lp rounding and greedy augmentation. 1.52 performance guarantee was obtained by Mahdin,Yeh, and Zhang [10] using dual fitting and greedy augmentation in n^3 running time, where n is number of vertex in given graph. An optimal bi factor approximation algorithm for the metric UFLP with 1.5 performance guarantee proposed by Byrka and Aardal [11] is the best known performance guarantee for this problem.
In this thesis we revisit Jain and Vazirani's algorithm but we use a different linear program for the problem in order to see if it leads to an improved performance.

2. UNCAPACITED FACILITY LOCATION PROBLEM

Let G be a bipartite graph with bipartition (F, D) where F vertices denote the facility locations and D vertices denote the cities or demand points. The edge weight c_{ij} for the edge (i, j) denote the cost of setting up a supply route from facility location i to the city j. We will denote $|F|$ and $|D|$ by n_f and n_d respectively. f_i denotes the cost of setting up the facility at location i.
The problem is to determine a subset $S \subseteq F$ where the facilities must be setup and an assignment $\phi : D \rightarrow S$ of demand points to facilities in S, so that $\sum_{i \in S} f_i + \sum_{j \in D} c_{\phi(j)j}$ is minimum.

3. JAIN AND VAZIRANI'S 3-APPROXIMATION SOLUTION FOR UNCAPACITATED FACILITY LOCATION PROBLEM

The problem is expressed as a linear program where y_i is an indicator variable denoting whether facility at i is open, and variable x_{ij} to indicate whether demand point j is served by facility i, integer programming formulation of this problem is as follows.

$$\begin{aligned} \text{minimize} \quad & \sum_{i\in F, j\in D} c_{ij}x_{ij} + \sum_{i\in F} f_i y_i \\ \text{subject to} \quad & \sum_{i\in F} x_{ij} \geq 1,\ j\in D, \\ & y_i - x_{ij} \geq 0,\ i\in F,\ j\in D, \\ & x_{ij} \in \{0,1\}\ i\in F,\ j\in D, \\ & y_i \in \{0,1\}\ i\in F. \end{aligned}$$

The first set of constraints ensures that each city is connected to at least one facility, and second ensures that demand points are served only by open facilities.
Consider the following LP-relaxation of UFLP.

$$\begin{aligned} \text{minimize} \quad & \sum_{i\in F, j\in D} c_{ij}x_{ij} + \sum_{i\in F} f_i y_i \\ \text{subject to} \quad & \sum_{i\in F} x_{ij} \geq 1,\ j\in D, \\ & y_i - x_{ij} \geq 0,\ i\in F,\ j\in D, \\ & x_{ij} \geq 0,\ i\in F,\ j\in D, \\ & y_i \geq 0,\ i\in F. \end{aligned}$$

The dual program of UFLP is formulated as follows.

$$\begin{aligned} \text{minimize} \quad & \sum_{j\in D} z_j \\ \text{subject to} \quad & z_j - p_{ij} \leq c_{ij},\ i\in F,\ j\in D, \\ & \sum_{j\in D} p_{ij} \leq f_i,\ i\in F, \\ & z_j \geq 0,\ j\in D, \\ & p_{ij} \geq 0,\ i\in F,\ j\in D. \end{aligned}$$

Let (x, y) and (z, p) be an optimal primal and dual solution respectively.
The primal and dual complementary slackness conditions are as follows.
Primal complementary slackness conditions:

$$\forall i\in F, j\in D: x_i > 0 \Rightarrow z_j - p_{ij} = c_{ij} \tag{1}$$

$$\forall i\in F: y_i > 0 \Rightarrow \sum_{j\in D} p_{ij} = f_i \tag{2}$$

Dual complementary slackness conditions:

$$\forall j\in D: z_j > 0 \Rightarrow \sum_{i\in F} x_{ij} = 1 \tag{3}$$

$$\forall i\in F, j\in D: p_{ij} > 0 \Rightarrow y_i = x_{ij} \tag{4}$$

Let i be a facility for which $y_i \geq 0$. Then $\sum_{j\in D} p_{ij} = f_i$ by equation 2. Thus we can interpret that demand point j is willing to pay amount p_{ij} to setup the facility at i. To open the facility at i, the demand points have to pay cost f_i. This interpretation can also be seen as follows. Let j be a demand point that is not served by i i.e., $x_{ij} = 0$. Since $x_{ij} \neq y_i$ from equation 4 implies that $p_{ij} = 0$, hence d_j does not contribute in setting up the facility at f_i, which it is not assigned to it.
To interpret equation 1 we define z_j to be the cost paid by the demand point d_j towards its share of setting up of the facility assigned to it and the cost of connecting d_j with that facility.

3.1 Primal-dual schema based algorithm

Their algorithm consists of two phases. In the first phase a primal/dual solution is computed using the complementary conditions. In the second phase the primal solution is refined.

3.1.1 Phase 1. In first phase they defined a notion of time so that each event can be associated with time. This phase starts with time $t = 0$ and each demand point is declared as unconnected at the beginning. Following process is performed after each unit of time. Dual variable z_j is raised by a unit after each unit of time elapses. For every j if there exists an i such that $z_j = c_{ij}$, the pair (i, j) is declared a tight edge. From now p_{ij} is also incremented by one unit after every unit of time so the first constraint of the dual program is never violated. Declare edge (i, j) as special edge which has $p_{ij} \geq 0$.
A facility i is said to be fully paid for if $\sum_{j\in D} p_{ij} \geq f_i$ and declare this facility temporarily open. All the demand points j are said to be tight which are connected to the facility i and declare i as connecting witness for those j. In future if any unconnected demand point j gets connected with i declare (i, j) a edge but not special because of $p_{ij} = 0$ and declare i as connecting witness for this demand point j. When all demand points get connected with some temporarily open facility, first phase terminates.

3.1.2 Phase 2. In phase-1 a demand point may get attached to more than one facility. Since each such point only needs to be attached to only one facility, we select one facility for each demand point in the second phase.
Let F_t denotes the set of temporarily open facilities and T denotes the subgraph of G comprising all special edges. Graph T^2 is defined on the same vertices but (u, v) is defined an edge iff there is a path between u and v in T having length at most 2. Let H be the subgraph of T^2 induced on F_t.
Compute a maximal independent set I in H. Declare the facilities in I to be permanently open. Each city will have at most one neighbor in I in H. If a city j has a neighbor i in I, then define $\phi(j) = i$. Otherwise select any vertex i' in F_t which is its connecting witness (which is ensured in the first phase). If $i' \in I$, then set $\phi(j) = i'$. Otherwise due to maximality of I as an independent set, there must be a neighbor of i' in I, say i''. Set $\phi(j) = i''$.

3.2 Analysis of the Algorithm

Dual variable z_j comprises the primal costs of connecting cities to facilities and the cost of opening the facilities. If j is connected in H to $\phi(j) = i$, then define $z_j^m = p_{ij}$ and $z_j^n = c_{ij}$. If j is not connected to $\phi(j) = i$, then $z_j^m = 0$ and $z_j^n = z_j$.

LEMMA 1. *Let $i \in I$ then,*
$\sum_{\phi(j)=i} z_j^m = f_i$.

PROOF. In phase-1 facility i was temporarily open and only special edges contributed to open it, i.e.,

$$\sum_{(i,j) \text{is special edge}} p_{ij} = f_i$$

□

COROLLARY 1. $\sum_{i\in I} f_i = \sum_{j\in D} z_j^m$

Note that only the directed connected demand points contributed to open facilities. Also if a city j was connected to the facility $i = \phi(j)$ in H, then $z_j^n = c_{ij}$. The only case that needs to be addressed is when j was not connected to i in H. In this case it was connected to some i' which was a neighbor of i. So we need to relate $c_{i'j}$ which is accounted for in z_j and $c_{i'j}$ which actually occurs in this solution.

LEMMA 2. $c_{ij} \leq 3z_j^m$, *for all indirectly connecting witness, where $\phi(j) = i$.*

PROOF. Since j is indirectly connected to i so there must be a tight edge (i', j) and an edge (i, i') in H. Since $H = T^2$, there must be a city j' such that (ij') and $(i'j')$ are both edges in T, i.e., both are special edges. Thus $z_{j'} \geq c_{ij'}$ and $z_{j'} \geq c_{i'j'}$. Since both edges are special, they must have gone tight before i and i' both were declared temporarily open. Let t_1, t_2 be times at which i and i' were declares temporarily open. So $z_{j'} \leq \min\{t_1, t_2\}$. On the other hand i' is the connecting witness of j so $z_j \geq t_2 \geq z_{j'}$. Putting these inequalities together we have $z_j \geq c_{ij'}$ and $z_j \geq c_{i'j'}$. From triangular inequality $c_{ij} \leq 3z_j = 3z_j^m$. □

We have the following result which establishes that this algorithm gives a 3 approximation.

THEOREM 1. *The primal and dual solutions constructed by the algorithm hold following inequality*
$\sum_{i\in F, j\in D} c_{ij}x_{ij} + 3\sum_{i\in F} F_i y_i \leq 3\sum_{j\in D} z_j$

This analysis is tight and it can be seen by the following example 1.

3.3 A Tight Example

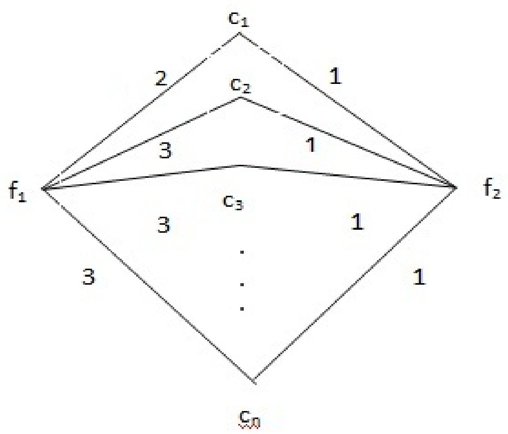

Fig. 1. Graph of Tight Example

The graph in figure1 has n cities $d_1, \ldots d_n$ and two facilities f_1 and f_2. Cost $c_{11} = 2$, $c_{1j} = 3$ for all $j > 1$, and $c_{2j} = 1$ for all $j \geq 1$. The opening cost of f_1 and f_2 are ϵ and $(n+1)\epsilon$, respectively, for a small positive $\epsilon \leq\leq 1$. The Jain-Vazirani algorithm computes a solution in which f_1 is open and all cities are served by it. This leads to a total cost of $3n - 2 + \epsilon$. It about three times the optimal cost $n + (n+1)\epsilon$ in which only f_2 must be open and all cities must be associated with it.

4. MOTIVATION

It is easy to see that in the above example the optimal solution opens the facility only at f_2. But the above algorithm fails to achieve the optimal solution even for such a simple situation. This shows that the algorithm needs to be improved.
In the following section we modify the algorithm by describing the conditions of the problem using an alternative linear program and again try to devise an algorithm using primal-dual technique.

5. PRIMAL-DUAL METHOD TO SOLVE UFLP WITH ALTERNATE LINEAR PROGRAM

We again attempt to solve this problem using primal-dual method but we will consider an alternative linear program. We will formulate linear program for uncapacitated facility location problem by first writing the LP for the capacitated facility location program and then take its restriction to uncapacitated case.

5.1 Capacitated Facility Location Problem (CFLP)

Let G be a bipartite graph with bipartition (F, D) where number of facility locations $n_f = |F|$, the number of demand points $n_d = |D|$, the costs of opening facilities f_i,$1 \leq i \leq n_f$, the demands d_j, $1 \leq j \leq n_d$ and the connecting costs c_{ij}, $1 \leq i \leq n_f, 1 \leq j \leq n_d$. A city j requires r_j supply and it may procure it from one or more locations. The supply from facility i to the city j is charged at the rate of c_{ij} per unit. Furthermore, if a facility is built at f_i, then it will have a capacity of s_i. It will be able to supply at most s_i units.
The solution includes a subset $S \subseteq F$ and an assignment $\alpha : S \times D \to \mathbb{R}$ where facilities will be open at all locations in S and $\alpha(i, j)$ (equivalently, α_{ij}) denotes the amount of product being supplied by facility i to the city j. The objective is to minimize $\sum_{i\in S}(f_i + \sum_{j\in D} c_{ij}\alpha_{ij})$ subject to the conditions that $\sum_j \alpha(i, j) \leq s_i$ for all $i \in S$ and $\sum_{i\in S}\alpha_{ij} \geq r_j$ for all $j \in D$.
Consider the following integer program for this problem. In this program, y_i is an indicator variable denoting whether facility i is selected for opening.

$$\begin{aligned}
\text{minimize} \quad & \sum_{i\in F, j\in D} c_{ij}\alpha_{ij} + \sum_{i\in F} f_i y_i \\
\text{subject to} \quad & \sum_{i\in F} \alpha_{ij} \geq r_j,\ j \in D, \\
& s_i y_i + \sum_{j\in D} -\alpha_{ij} \geq 0,\ i \in F, \\
& \alpha_{ij} \geq 0,\ i \in F,\ j \in D, \\
& y_i \in \{0, 1\},\ i \in F.
\end{aligned}$$

LP-relaxation of CFLP is as follows.

$$\begin{aligned}
\text{minimize} \quad & \sum_{i\in F, j\in D} c_{ij}\alpha_{ij} + \sum_{i\in F} f_i y_i \\
\text{subject to} \quad & \sum_{i\in F} \alpha_{ij} \geq r_j,\ j \in D, \\
& s_i y_i + \sum_{j\in D} -\alpha_{ij} \geq 0,\ i \in F, \\
& \alpha_{ij} \geq 0,\ i \in F,\ j \in D, \\
& y_i \geq 0,\ i \in F.
\end{aligned}$$

The dual program for CFLP is as follows.

$$\begin{aligned}
\text{minimize} \quad & \sum_{j\in D} z_j \\
\text{subject to} \quad & z_j - p_i \leq c_{ij},\ i \in F,\ j \in D, \\
& s_i p_i \leq f_i,\ i \in F, \\
& z_j \geq 0,\ j \in D, \\
& p_i \geq 0,\ i \in F.
\end{aligned}$$

The dual variables may be interpreted as follows. z_j represents the total cost incurred by city j and p_i denotes the setup cost per unit supply.
Primal complementary slackness conditions:

$$\forall i \in F : y_i > 0 \Rightarrow p_i s_i = f_i \quad (5)$$

$$\forall i \in F, j \in D : \alpha_{ij} > 0 \Rightarrow z_j - p_i = c_{ij} \quad (6)$$

Dual complementary slackness conditions:

$$\forall i \in F : p_i > 0 \Rightarrow s_i y_i - \sum_{j\in D} \alpha_{ij} = 0 \quad (7)$$

$$\forall j \in D : z_j > 0 \Rightarrow \sum_{i\in F} \alpha_{ij} = r_j \quad (8)$$

For a k-approximation we relax these conditions as follows:
Relaxed primal complementary slackness conditions:

$$\forall i \in F : y_i > 0 \Rightarrow f_i/(ks_i) \leq p_i \leq f_i/s_i \tag{9}$$

$$\forall i \in F, j \in D : \alpha_{ij} > 0 \Rightarrow c_{ij}/k \leq z_j - p_i \leq c_{ij} \tag{10}$$

Relaxed dual complementary slackness conditions:

$$\forall i \in F : p_i > 0 \Rightarrow s_i y_i - \sum_{j \in D} \alpha_{ij} \geq 0 \tag{11}$$

$$\forall j \in D : z_j > 0 \Rightarrow \sum_{i \in F} \alpha_{ij} = r_j \tag{12}$$

Algorithm 1 is the heuristic algorithm for capacitated case.

Algorithm 1 Primal-Dual Algorithm for a given k

Initialize all dual variables to zero;
Initialize an empty graph H over vertices $D \cup F$;
Unlock all z_j's and lock all p_i's;
repeat
  Raise all unlocked dual variables with time till one of the following events occur (i) $c_{ij}/k = z_j$ for some i, j or (ii) $z_j = c_{ij} - p_i$ for some i or (iii) $p_i = f_i/s_i$;
  if for each i, j such that $c_{ij}/k = z_j$ is just realized **then**
    Set $p_i = f_i/(ks_i)$;
  end if
  if For each i, j such that $c_{ij} = z_j - p_i$ is just realized **then**
    Add the edge (i, j) in H;
  end if
  if $p_i < f_i/s_i$ **then**
    Unlock p_i;
  end if
  if $p_i = f_i/s_i$ is just realized **then**
    Lock p_i;
  end if
  if $c_{ij} = z_j - p_i$ **then**
    Lock z_j;
  end if
  if Graph changes from the previous iteration **then**
    Solve the following LP:$s_i y_i - \sum_j \alpha_{ij} \geq 0$ for all i with degree greater than zero, $\sum_i \alpha_{ij} = r_j$ for feasibility;
  end if
  if Feasible **then**
    Report solution;
  else if H is a complete bipartite graph **then**
    Report no solution and stop.
  end if
until A solution is realized or no solution is found feasible

Due to the demand and supply parameters r_j and s_i, we find it difficult to determine a value of k for which a solution can be guaranteed. Hence we propose this solution as a heuristic approach. For an instance of a problem one can run this algorithm for increasing values of k till a solution is found.

5.2 Adopting the algorithm to uncapacitated case

In order to specialize the algorithm for capacitated case to uncapacitated case we need to reinterpret s_i's and r_j's. Since in the latter case only one connection from each city to a facility is required, we may define $r_j = \delta \leq\leq 1$.
To define s_i recall the primal condition $s_i y_i \geq \sum_j \alpha_{ij}$. To enable the case when all cities choose to connect to the same facility, we define each s_i equal to $|D| = n_d$.
The LP in the above algorithm is no longer required in the uncapacitatted case. In stead, it is sufficient to check if each city is connected to at least one facility. It is possible that a city may get connected to more than one facility. Hence we need to prune out unnecessary connections. Hence we randomly remove edges until the degrees all cities reduce to one.
Algorithm 2 is the heuristic algorithm for uncapacitated case.

Algorithm 2 Primal-Dual Algorithm for a given k

Initialize all dual variables to zero;
Initialize an empty graph H over vertices $D \cup F$;
Unlock all z_j's and lock all p_i's;
repeat
  Raise all unlocked dual variables with time till one of the following events occur (i) $c_{ij}/k = z_j$ for some i, j or (ii) $z_j = c_{ij} - p_i$ for some i or (iii) $p_i = f_i/n_d$;
  if for each i, j such that $c_{ij}/k = z_j$ is just realized **then**
    Set $p_i = f_i/(kn_d)$;
  end if
  if For each i, j such that $c_{ij} = z_j - p_i$ is just realized **then**
    Add the edge (i, j) in H;
  end if
  if $p_i < f_i/n_d$ **then**
    Unlock p_i;
  end if
  if $p_i = f_i/n_d$ is just realized **then**
    Lock p_i;
  end if
  if $c_{ij} = z_j - p_i$ **then**
    Lock z_j;
  end if
  if All cities are connected to at least one facility **then**
    repeat
      Randomly prune edges;
    until The degree of all cities reduce to one
  end if
  Return H and stop.
until A solution is realized

In order to test this algorithm we ran it on the tight example. It resulted into a graph in which each city was connected to f_2. We further ran the algorithm on the same example but took ϵ to very very large. In that case all cities were connected to f_1. In each case, the solutions were optimal.

6. CONCLUSION

We revisited the uncapacitated facility location problem and its algorithm given by Jain and Vazirani. We gave a heuristic algorithm derived from primal-dual schema which perform better on a tight example compare to performance of Jain and Vazirani.

7. REFERENCES

[1] J.Stollsteimer, "A working model for plant numbers and locations," *Journal of Farm Economics*, vol. 45, pp. 631–645, 1963.

[2] A. A. Kuehn and M. J. Hamburger, "A heuristic program for locating warehouse," *Management Science*, vol. 9, pp. 643–666, 1963.

[3] A. Manne, "A plant location and economy of scale decentralization and computation," *Management Science*, vol. 11, pp. 213–235, 1964.

[4] G.Cornuejola, G.L.Nemhauser, and L. Wolsey, "The uncapacitated facility location problem," *In P.Mirchandani and R.Francis editors Discrete Location Theory John Wiley and Sons New York*, pp. 119–171, 1990.

[5] G.Guha and Khullar, "Greedy strikes back:improved facility location algorithms," *Journal of Algorithms*, vol. 31(1), pp. 228–248, 1990.

[6] M.Sviridenko, "Approximation algorithms for facility location problem," *PhD thesis Stanford*, 2000.

[7] D.Hochbaum, "Heuristics for the fixed cost median problem," *Mathematical Programming*, vol. 22), pp. 148–162, 1982.

[8] D.B.Shmoys, E.Tardos, and K.I.Aardal, "Approximation algorithms for facility location problems," *In proceedings of the* 29^{th} *Annual Symposium on Theory of Computing*, pp. 265–274, 1997.

[9] K. Jain and V. V. Vazirani, "Approximation algorithm for metric facility location and k-median problems using primal-dual schema and lagrangian relaxation," *ACM*, vol. 48, March 2001.

[10] M.Mahdian, Y.Ye, and J.zhang, "A 1.52 approximation algorithm for the uncapacitated facility location problem," *Manuscript*.

[11] J. Byrka, "An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem," *Proceedings of the* 10^{th} *International Workshop on Approximation and the* 11^{th} *International Workshop on Randomization and combinatorial Optimization Algorithm and Techniques Berlin Heidelberg Springer -Verlag*, pp. 29–43, 2007.