An Approximation Algorithm for the k-level Uncapacitated Facility Location Problem with Penalties

Mohsen Asadi[1], Ali Niknafs[1], Mohammad Ghodsi[2,3]

[1] Computer Engineering Department, Sharif University of Technology, Tehran, Iran
{mo_asadi, niknafs}@ce.sharif.edu

[2] Computer Engineering Department, Sharif University of Technology, Tehran, Iran
School of Computer Science, Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran
ghodsi@sharif.edu

Abstract. The k-level Uncapacitated Facility Location (UFL) problem is a generalization of the UFL and the k-median problems. A significant shortcoming of the classical UFL problem is that often a few very distant customers, known as outliers, can leave an undesirable effect on the final solution. This deficiency is considered in a new variant called UFL with outliers, in which, in contrast to the other problems that need all of the customers to be serviced, there is no need to service the entire set of customers. UFL with Penalties (UFLWP) is a variant of the UFL with outliers problem in which we will decide on whether to provide service for each customer and pay the connection cost, or to reject it and pay the penalty. In this paper we will propose a new 4-approximation algorithm for the UFLWP which is the first algorithm for this kind of problem.

Keywords: Approximation algorithms, Facility location problem, k-level facility location, Outliers

1. Introduction

Facility location problem is a fundamental problem in theoretical science. In the classical single-level Uncapacitated Facility Location (UFL) problem, we are given two sets: one set for facilities (F), and another one for customers (C). There is a specified *connection* cost $c_{ij} \geq 0$ between every pair $i, j \in F \cup C$. Opening a facility $i \in F$ causes a fixed nonnegative *open* cost f_i. The goal of this problem is to locate facilities and assign each customer to one of the opened facilities, such that the total cost of opening facilities and of servicing the customers would be minimized. That is, in a formal manner, we should identify a subset of facilities $S \subseteq F$ and to serve the customers in C by the facilities in S, so as to minimize the total cost function:

[3] This author's work was in part supported by a grant from IPM (N. CS2386-2-01.)

$$Cost(S) = \sum_{i \in S} f_i + \sum_{\substack{i \in S \\ j \in C}} c_{ij}$$

In this paper, we adopt notions of two variants of uncapacitated facility location problem and mix them to form a new problem, known as *k-level Uncapacitated Facility Location with Outliers* problem. The problem of the metric k-level UFL problem is a generalization of the UFL and the k-median problems. In this problem, each customer must be serviced by a sequence of k different kinds of facilities located in k levels of hierarchy. This well-known problem has various applications which can be found in [1].

Another variant of the facility location problem is called UFL with outliers. In this kind of problem, in contrast to the forementioned other problems in which all of the customers must be serviced, there is no need to service the entire set of customers. This is due to the fact that often a few very distant customers, called *outliers*, can leave an undesirable effect on the result solution. This deficiency is considered in a new variant called UFL with outliers, in which very distant outlier may be ignored due to economical issues [2]. Two variants of the UFL with outliers problem have been proposed in the literature:

- *Uncapacitated Facility Location with Penalties (UFLWP)*: in this problem a penalty r_j is assigned to each customer$j$. For each customer we may decide to either provide service, and pay the connection cost to its nearest facility or to ignore it and pay the penalty. Setting the penalties to ∞ gives the standard formulation.
- *Robust Facility Location (RFL)*: in this problem we are given a parameter η. The problem is to locate facilities so as to minimize the service cost to any subset of facilities of size at least η. In case that $\eta = n$ the problem is equivalent to the standard formulation.

See [3] for a comprehensive review on other variants of the facility location problem with outliers. A number of efficient approaches have been proposed in recent years, which can be roughly classified into several categories: greedy approach [4], LP rounding techniques [5], local search heuristics [6], primal-dual method [5], game theory [7], and randomization technique [8]. In some sense, techniques from different categories complement each other, and could be combined to achieve improved approximation algorithms [4]. The first approximation algorithm for metric k-level UFL problem proposed in [9] uses the primal-dual scheme in linear programming which has an approximation factor of 6. Another algorithm using game theory achieves the same approximation factor [7]. The approximation factor is improved to 4 using the greedy approach [10]. Aardal et. al. [8] proposes a 3-approximation algorithm using linear programming relaxation.

For UFL with outliers problem three constant approximation algorithms have been proposed [11, 5, 9]. A 3-approximation algorithm by means of primal-dual scheme has been suggested for both kinds of UFL with outliers problem [11]. For the UFLWP problem, an algorithm using the LP rounding technique has been proposed in [5] with $2 + 2/e$ approximation factor, where e is the natural logarithmic base, whereas a 2-approximation algorithm is proposed in [4].

In this paper we examine the metric k-level uncapacitated facility location problem with outliers. This problem is considered as an extension to the classical k-level UFL problem in which the outliers are ignored to achieve an improved level of service to majority of customers. We adopt the UFLWP variant of the single-level UFL with outliers problem [5], and propose new algorithms for the case with k-levels of facilities by applying the LP techniques. To our knowledge, there is no algorithm proposed for k-level UFL with penalties.

The paper is organized as follows. Next section includes a brief description of the k-level UFL with penalties. Section 3 presents our proposed 4-approximation algorithm for the UFLWP problem.

2. Problem description

In this section, first we present a formal and precise definition of the k-level UFL problem. Then we explain the required changes need to be performed to obtain the k-level version of UFLWP problem. Finally we formulate the UFLWP by taking the advantage of linear programming technique.

2.1. k-level UFL

Let C be the set of customers. Each customer $j \in C$ must be assigned to precisely one facility at each of the k levels. Let F^l be the set of locations where facilities on level l, $1 \leq l \leq k$, may be located and assume that the sets $F^l$are pairwise disjoint. $F = \cup_{l=1}^{k} F^l$ is considered as the set of all such locations. The cost of setting up a facility at location i is f_i, which $f_i > 0$ for each $i \in F$. The cost of connection between points $i, j \in F \cup C$ is equal to c_{ij}, which $c_{ij} > 0$. Throughout this paper, we make the metric assumption of the k-level UFL problem, i.e. for each $i, j, k \in F \cup C$ we have $c_{ik} \leq c_{ij} + c_{jk}$, which satisfies the triangle inequality. We shall use p to denote a sequence of facilities $i_l \in F^l, l = 1, \dots, k$, and shall refer to p as a *path* of facilities. The set of all possible paths is denoted by P. each customer must be assigned to precisely one path $p \in P$. The total connection cost incurred by assigning customer j to path $p = (i_1, i_2, \dots, i_k)$ is equal to $c_{jp} = c_{i_1,i_2} + c_{i_2,i_3} + \cdots + c_{i_{k-1},i_k} + c_{i_k,j}$.

The goal of this problem is to assign each customer to a sequence of k facilities, one at each level, such that satisfying the service demand of each customer and minimize the total cost of opening facilities and connection costs.

2.2. k-level UFLWP

This problem is derived from the classical k-level UFL problem. Each customer $j \in C$ will be either serviced or rejected completely. If the customer is planned to be serviced then it should be assigned to a sequence of facilities one in each level. An r_j

parameter is assigned to each customer $j \in C$ indicating the penalty of rejecting a customer. The goal of this problem is to find:

1. A subset $Q \subseteq C$ of customers whose demands should be rejected,
2. A subset $S \subseteq F$ as the locations for opening facilities, such that $S = \cup_{l=1}^{k} S^l$, and
3. Remaining customers are assigned to a sequence of k facilities, one at each level, so that each customer will be serviced.

The following is a formulation for the k-level UFLWP by means of integer programming (denoted as IP_1). Let suppose x_{jp} equals one, if customer j is assigned to path p, otherwise it equals zero. In addition, let y_{i_l} equals 1 if the facility $i_l$on level l is open. Furthermore z_j equals 1 if the customer is rejected and 0 if the customer is serviced. Equations (1)-(6) represent the integer program of this problem.

$$minimize$$

$$Z_{IP_1} = \sum_{l=1}^{k} \sum_{i_l \in F^l} f_{i_l} y_{i_l} + \sum_{p \in P} \sum_{j \in C} c_{jp} x_{jp} + \sum_{j \in C} r_j z_j \quad (1)$$

$$subject\ to$$

$$z_j + \sum_{p \in P} x_{jp} \geq 1 \quad for\ each\ j \in C \quad (2)$$

$$\sum_{P: p \ni i_l} x_{jp} - y_{i_l} \leq 0$$

$$for\ each\ j \in C\ and\ i_l \in F^l, l = 1, \dots, k \quad (3)$$

$$x_{jp} \in \{0,1\}$$

$$for\ each\ j \in C\ and\ i_l \in F^l, l = 1, \dots, k \quad (4)$$

$$y_{i_l} \in \{0,1\} \quad for\ each\ i_l \in F^l, l = 1, \dots, k \quad (5)$$

$$z_j \in \{0,1\} \quad for\ each\ j \in C \quad (6)$$

3. Algorithm

In this section, we present a constant factor approximation algorithm for the k-level UFLWP problem, by adopting the *LP rounding technique*. The linear relaxation version of $IP_1$is obtained by relaxing the constraints (4), (5), and (6). That is these constraints are replaced with inequalities $x_{jp} \geq 0, y_{i_l} \geq 0,$ and $z_i \geq 0$ respectively. Our solution to the UFLWP problem, first solves the linear relaxation program for each $j \in C$. Let LP_1 denote this relaxation program, which will be solved by means of existing LP solving techniques. The obtained solution would be fractional; however, we need to round the solution to achieve a near optimum integral solution suitable for the actual integer program. Formulation of the LP_1 relaxation program is as follows:

$$minimize$$

$$Z_{LP_1} = \sum_{l=1}^{k} \sum_{i_l \in F^l} f_{i_l} y_{i_l} + \sum_{p \in P} \sum_{j \in C} c_{jp} x_{jp} + \sum_{j \in C} r_j z_j \quad (7)$$

$$subject\ to$$

$$z_j + \sum_{p \in P} x_{jp} \geq 1 \ \ for\ each\ j \in C \quad (8)$$

$$\sum_{P:p\ni i_l} x_{jp} - y_{i_l} \le 0 \quad for\ each\ j \in C\ and\ i_l \in F^l, l = 1, \dots, k \quad (9)$$

$$\sum z_j \le l \quad for\ each\ j \in C \quad (10)$$

$$x_{jp} \ge 0 \quad for\ each\ j \in C\ and\ p \in P \quad (11)$$

$$y_{i_l} \ge 0 \quad for\ each\ i_l \in F^l, l = 1, \dots, k \quad (12)$$

$$z_j \ge 0 \quad for\ each\ j \in C \quad (13)$$

Note that in the above program, due to the constraints (8) and (11), there is no need to include the upper bound $x_{jp} \le 1$.

Let (x', y', z')be the solution of LP_1, and w'be the optimal objective value of it. Likewise, suppose that OPT be the optimal objective value of IP_1. Obviously, $OPT \ge w'$.

After solving the linear program and obtaining the above values, the customers whom should be rejected are then identified. This is done by comparing the value of $z'_j$of each customer $j \in C$ with a threshold. Let $\delta \in (0,1)$ be this threshold. Moreover, let $Q_s$be the set of rejected customers, i.e. $Q_s = \{j \in C \mid z'_j \ge \delta\}$. If $Q_s = C$, then every customer must be rejected, so the algorithm is terminated. Otherwise, the remaining customers need to be serviced via opening new facilities.

In order to define an upper bound for the obtained solution, we introduce auxiliary variables $\bar{x}$ and $\bar{y}$. It is shown in the Lemma 1 how they can be used to analyze the approximation factor of the algorithm. More specifically, two auxiliary variables are defined as follows:

$$\bar{x}_{jp} = \begin{cases} 0 & if\ j \in Q_s \\ \min\left(\frac{1}{1-\delta}\, x'_{jp}, 1\right) & if\ j \notin Q_s \end{cases} \quad (14)$$

$$\bar{y}_{i_l} = \min\left(\frac{1}{1-\delta} y'_{i_l}, 1\right) \quad for\ each\ i_l \in F^l, l = 0, \dots, k \quad (15)$$

To find a solution with the minimum cost of opening facilities and of assigning customers which are selected to be serviced, i.e. those who are not placed in Q_s, we introduce the following integer program (denoted as IP_2) for the new UFL problem with F as the set of facilities and $\mathcal{C} = \{j \in C \mid j \notin Q_s\}$ as the set of customers. After relaxing the constraints,

$$x_{jp} \in \{0,1\} \quad for\ each\ j \in \mathcal{C}\ and\ i_l \in F^l, l = 1, \dots, k$$

$$y_{i_l} \in \{0,1\} \quad for\ each\ i_l \in F^l, l = 1, \dots, k$$

the linear relaxation of IP_2 is obtained (denoted as LP_2), which is presented as follows:

$$minimize$$

$$Z_{LP_2} = min \sum_{l=1}^{k} \sum_{i_l \in F^l} f_{i_l} y_{i_l} + \sum_{p \in P} \sum_{j \in \mathcal{C}} c_{jp} x_{jp} \quad (16)$$

$$subject\ to$$

$$\sum_{p \in P} x_{jp} \ge 1 \quad for\ each\ j \in \mathcal{C} \quad (17)$$

$$\sum_{P:p\ni i_l} x_{jp} - y_{i_l} \le 0$$

$$for\ each\ j \in \mathcal{C}\ and\ i_l \in F^l, l = 1, \dots, k \quad (18)$$
$$x_{jp} \geq 0 \quad for\ each\ j \in \mathcal{C}\ and\ p \in P \quad (19)$$
$$y_{i_l} \geq 0 \quad for\ each\ i_l \in F^l, l = 1, \dots, k \quad (20)$$

Let w^* be the optimal objective value of LP_2 and let A be an approximation algorithm to solve $IP_2$for the k-level UFL problem (such as [8]). Approximation factor of A would be $\lambda = \frac{A^*}{w^*}$, where $A^*$is the objective value of the integral solution returned by algorithm A. Since, $OPT_A > w^*$, then the approximation factor of algorithm A, i.e. $\frac{A^*}{OPT_A}$, will be no larger than λ. Therefore, the total cost of opening facilities and connections between customers and opened facilities, returned by algorithm A is bounded by λw^*.

By applying algorithm A to IP_2, customers belonging to $\mathcal{C}$ are assigned to facilities in F. Afterwards, the obtained solution is combined with the set of rejected customers (those belonging to Q_s) to form a solution to IP_1. The total cost of the latest obtained solution is no more than $\lambda w^* + \sum_{j \in Q_s} r_j$.
In the following describe the algorithm and prove its approximation factor.

Algorithm 1 (Metric *k*-level UFLWP – factor 4)
1. Solve the linear program $LP_1$for facility set F and customer set C.
2. Find the set of customers to be rejected (Q_s), and reject them, if all customers are rejected then terminate, Define $\mathcal{C} = C \backslash Q_s$.
3. Apply approximation algorithm A to the problem with F as facility set and $\mathcal{C}$as customer set. The algorithm opens facilities in F and assigns customers belonging to $\mathcal{C}$ according to the solution returned by A.

To determine the approximation factor we need to find the relation between the solution obtained for $IP_1$and IP_2. This relationship can be identified by means of the next two lemmas.

Lemma 1. Let $(\bar{x}, \bar{y})$ be a feasible solution for LP_2 and,

$$w^* \leq \frac{1}{1-\delta} \sum_{l=1}^{k} \sum_{i_l \in F^l} f_{i_l} y'_{i_l} + \frac{1}{1-\delta} \sum_{p \in P} \sum_{j \in C} c_{jp} x'_{jp} \quad (21)$$

Proof. First, we need to show that $(\bar{x}, \bar{y})$ is a feasible solution for LP_2. We know that $\bar{x}_i$ and $\bar{y}_i$ always have positive values. So, we only need to show that constraints (17) and (18) are correct.
Suppose an arbitrary customer $j \notin Q_s$, we have:

$$\sum_{p \in P} \bar{x}_{jp} = \sum_{p \in P} \min\left(\frac{1}{1-\delta} x'_{jp}, 1\right) \quad (22)$$

If there exists a path $p \in P$, such that $\frac{1}{1-\delta} \bar{x}_{jp} \geq 1$, the constraint (17) would be obviously correct. Otherwise,

$$\sum_{p \in P} \min\left(\frac{1}{1-\delta} x'_{jp}, 1\right) = \sum_{p \in P} \frac{1}{1-\delta} x'_{jp} \geq \frac{1}{1-\delta}\left(1 - z'_j\right) \quad (23)$$

The above inequality is derived from constraint (2) and the fact that (x', y', z') is feasible solution for LP_1. Since we have $j \notin Q_s$, then $z_j' \leq \delta$. Therefore,

$$\sum_{p\in P} \min\left(\frac{1}{1-\delta}\, x'_{jp}, 1\right) \geq \frac{1}{1-\delta}\left(1 - z'_j\right) \geq 1 \quad (24)$$

and so the constraint (17) would be correct.

To prove the correctness of the constraint (18), we have $x'_{jp} \leq y'_{i_l}$ due to the constraint (3), so,

$$\bar{x}_{jp} = \min\left(\frac{1}{1-\delta}x'_{jp}, 1\right) \leq \min\left(\frac{1}{1-\delta}y'_{i_l}, 1\right) = \bar{y}_{i_l} \quad (25)$$

Now we have to bound the value of w^*. Since, $(\bar{x}, \bar{y})$ is a feasible solution for LP_2, and $w^*$is the optimal solution for LP_2, then w^* would be the lower bound of the objective value of $(\bar{x}, \bar{y})$. Thus,

$$w^* \leq \sum_{l=1}^{k} \sum_{i_l \in F^l} f_{i_l}\bar{y}_{il} + \sum_{p\in P}\sum_{j\in C} c_{jp}\bar{x}_{jp} \quad (26)$$

By the definitions of $\bar{x}$ and $\bar{y}$, we have:

$$\begin{aligned} w^* &\leq \sum_{l=1}^{k} \sum_{i_l \in F^l} \frac{1}{1-\delta} f_{i_l} y'_{i_l} + \sum_{p\in P}\sum_{j\in C} \frac{1}{1-\delta} c_{jp} x'_{jp} \\ &\leq \sum_{l=1}^{k} \sum_{i_l \in F^l} \frac{1}{1-\delta} f_{i_l} y'_{i_l} + \sum_{p\in P}\sum_{j\in C} \frac{1}{1-\delta} c_{jp} x'_{jp} \end{aligned} \quad (27)$$

□

Lemma 2. Suppose that A is the approximation algorithm with factor λ. Then the above algorithm for the metric k-level UFLWP problem is a $(1+\lambda)$-approximation algorithm.

Proof. As we stated before the cost of algorithm is:

$$\begin{aligned} &\lambda w^* + \sum_{j\in Q_s} r_j \\ &\leq \frac{\lambda}{1-\delta}\left(\sum_{l=1}^{k} \sum_{i_l \in F^l} f_{i_l} y'_{i_l} + \sum_{p\in P}\sum_{j\in C} c_{jp} x'_{jp}\right) + \sum_{j\in Q_s} r_j \\ &\leq \frac{\lambda}{1-\delta}\left(\sum_{l=1}^{k} \sum_{i_l \in F^l} f_{i_l} y'_{i_l} + \sum_{p\in P}\sum_{j\in C} c_{jp} x'_{jp}\right) + \sum_{j\in Q_s} \frac{1}{\delta} r_j\, z'_j \\ &\leq \max\left(\frac{\lambda}{1-\delta}, \frac{1}{\delta}\right) w' \\ &\leq \max\left(\frac{\lambda}{1-\delta}, \frac{1}{\delta}\right) OPT. \end{aligned} \quad (28)$$

If we take $\delta = \frac{1}{1+\lambda}$,the approximation algorithm would be $1+\lambda$.

□

Theorem 1. There is a polynomial time algorithm with approximation factor of 4 for the k-level UFLWP.

Proof. Since the algorithm presented in [1], which is supposed the algorithm A with approximation factor 3, so we can use it in step 3 of our algorithm. Therefore, due to lemma 2, the approximation factor of our algorithm is 4 for solving the k-level UFLWP.

□

4. Discussion

In this paper, we adopt notions of two variants of the uncapacitated facility location problem and mix them to form a new problem, known as k-level uncapacitated facility location with outlier. Afterwards, we represent the algorithm for the k-level UFLWP variant of it using LP technique. The approximation factor of the algorithm is 4. To our knowledge this is the only existing algorithm for this problem.

References

1. Sahin, G., Sural, H.: A review of hierarchical facility location models. J. Comput. Oper. Res. 34 (8), 2310--2331 (2007)
2. Arya, V., Garg, N., Khandekar, R., Munagala, K., Pandit, V.: Local search heuristic for k-median and facility location problems. In: Proceedings of the 33rd Annual ACM Symposium on theory of Computing (STOC '01), Greece, pp.21--29. ACM Press (2001)
3. Shmoys, D., Approximation algorithms for facility location problems. In: Proceedings of the 3rd International Workshop on Approximation Algorithms for Combinatorial Optimization. LNCS, vol. 1913, pp.27--33, Springer (2000)
4. Jain, K., Mahdian, M., Markakis, E., Saberi, A., Vazirani, V.: Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. J. ACM 50 (6), 795--824 (2003)
5. Xu, G., Xu, J.: An LP rounding algorithm for approximating uncapacitated facility location problem with penalties. J. Inf. Process. Lett. (94), pp.119--123 (2005)
6. Korupolu, M. R., Plaxton, C. G., Rajaraman, R.: Analysis of a local search heuristic for facility location problems. In: Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1--10 (1998)
7. Xu, D., Du, D.: The k-level facility location game. J. Oper. Res. Lett. 34 (4), 421--426 (2006)
8. Aardal, K., Chudak, F. A., Shmoys, D. B.: A 3-approximation algorithm for the k-level uncapacitated facility location problem. J. Inf. Process. Lett. 72 (5-6), 161--167 (1999)
9. Jain, K., Vazirani, V., Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. J. ACM (48), 274--296 (2001)
10. Zhang, P.: A new approximation algorithm for the k-facility Location Problem. TAMC'06, LNCS, vol. 3959, pp. 217--230. Springer-Verlag (2006)
11. Charikar, M., Khuller, S., Mount, D., Narasimhan, G.: Algorithms for facility location problems with outliers. In: Proceedings of Symposium on Discrete Algorithms, pp. 642--651 (2001)