Approximation Algorithms for Fair Range Clustering

Sedjro S. Hotegni, Sepideh Mahabadi, & Ali Vakilian
ICML, 2023

Jinwon Park, Jihu Lee
January 19, 2024

Seoul National University

Abstract

Fair range clustering problem

- Suppose data points are from ℓ different demographic groups
- Target is to pick k centers with the minimum ℓ_p-clustering cost
- Each group is at least *minimally represented* in the centers set and *no group dominates* the centers set

This paper provides an efficient constant factor approximation algorithm for the fair range ℓ_p-clustering

Main contribution

- finding an approximate functional solution (x, y) for an LP-relaxation of the Fair range clustering problem
- rounding the fractional solution to an integral solution with $e^{O(p)}$-approximation

Notations and preliminaries

- set of n points are given in a metric space (P, d)
- each points belongs to one of the ℓ disjoint demographics $P = P_1 \uplus P_2 \uplus \cdots \uplus P_\ell$
- $D \subseteq P$: set of clients
- $F \subseteq P$: set of facilities
- $[\alpha_i, \beta_i]$: interval for a number of centers for each group $i \in \ell$ i.e. $|D \cap P_i| \in [\alpha_i, \beta_i]$ for $i \in \ell$

Constant Factor Approximation Algorithm

K-center problem can be stated as follow;

$$\min \sum_{v \in D, u \in P} w_v c_{vu} x_{vu}$$

subject to

$$\sum_{u \in F} x_{vu} = 1 \qquad \text{for each } v \in D$$

$$x_{vu} \leq y_v \qquad \text{for each } v \in D, u \in F$$

$$\sum_{v \in D} y_v \leq k$$

$$x_{vu} \in \{0,1\} \qquad \text{for each } v \in D, u \in F$$

$$y_v \in \{0,1\} \qquad \text{for each } u \in F$$

where y_v indicates the location v is selected as a center and x_{vu} indicates if location u is assigned to the center at v.

Here demand w_v specifies the number of clients present at that location $u \in N$

Constant Factor Approximation Algorithm

K-center problem is a NP hard problem.

(Charikar et al, 2002) suggests linear programming relaxation to the integer program by replacing 0-1 constraints as

$$x_{vu} \geq 0 \qquad \text{for each } v \in D, u \in F$$

$$y_v \geq 0 \qquad \text{for each } u \in F$$

Fair Range LP

Theorem

1.1 For all $p \in [1, \infty)$, there exists a constant factor approximation algorithm for fair range k-clustering with the ℓ_p-objective that runs in polynomial time

Fair Range LP

LP-relaxation of Fair range clustering algorithm

$$\min \sum_{v \in D, u \in F} w(v) \cdot d(v,u)^p x_{vu}$$

subject to

$$\begin{aligned}
\sum_{u \in F} x_{vu} &\geq 1 && \forall v \in D \\
\alpha_i \leq \sum_{u \in P_i} y_u &\leq \beta_i && \forall i \in [\ell] \\
\sum_{u \in F} y_u &\leq k && \\
0 \leq x_{vu} &\leq y_u && \forall v \in D, u \in F
\end{aligned}$$

Rounding Algorithm

- For rounding, the paper proposes a new LP relaxation called Structured LP which is a simplification of Fair Range LP via theorem above.
- Structured LP is useful for rounding since the polyhedron constructed by the constraints of Structured LP is half-integral.
- A solution to an Structured LP has value either 0, 1/2 or 1.

Rounding Algorithm

Structured LP

$$\min \sum_{v \in D'} w'(v) \cdot \Delta(v)$$

such that

$$\alpha_i \leq \sum_{u \in F_i} y_u \leq \beta_i \qquad \forall i \in [\ell]$$

$$\sum_{u \in F} y_u \leq k$$

$$\sum_{u \in \mathcal{B}(v)} y_u \geq 1/2 \qquad \forall v \in D'$$

$$\sum_{u \in \mathcal{P}(v)} y_u \leq 1 \qquad \forall v \in D'$$

$$y_u \geq 0 \qquad \forall u \in F$$

Rounding Algorithm

- $\mathcal{R}(v) := \left(\sum_{u \in P} x^*_{vu} \cdot d(v,u)^p\right)^{1/p}$ is the fractional distance of a unit of demand at location v w.r.t the optimal solution (x^*, y^*)
- $\mathcal{B}(v) := \{u \in F | d(v,u) \leq 2^{1/p} \cdot \mathcal{R}(v)\}$ is the set of facilities at distance at most $2^{1/p} \cdot \mathcal{R}(v)$ from v
- $\mathcal{P}(v)$ is a super ball of v that consists of $\mathcal{B}(v)$ and a set of private facilities of v
- $\Delta(v) := d(v,v')^p + \sum_{u \in \mathcal{P}(v)} \left(d(v,u)^p - d(v,v')^p\right)$ is the minimum distance of v to facilites

Rounding Algorithm

Theorem

Given an instance (D, w) of fair range clustering with ℓ_p-cost and an optimal fractional solution (x, y) of Fair Range LP(D, w) with cost OPT_D, there exists a polynomial time algorithm that returns a set of locations $D' \subset D$ and a demand function $w' : D' \rightarrow \mathbb{R}$ such that

1. *For every pair for v_i, v_j in D', $d(v_i, v_j) \geq 2^{1+1/p} \max\{\mathcal{R}(v_i), \mathcal{R}(v_j)\}$*
2. *(x, y) is a feasible solution of Fair Range LP(D', w') of cost at most OPT_D*
3. *Any integral solution C of Fair Range LP(D', w') of cost z, can be converted in polynomial time to a feasible solution of Fair Range LP(D, w) of cost at most $4^p \cdot OPT_D + 2^{p-1} \cdot z$*

Rounding Algorithm

Theorem

There exists a polynomial time algorithm that outputs a fractional solution (x, y) of Fair Range LP(D', w') of cost $9^p \cdot OPT_D$, where OPT_D is the cost of an optimal solution of Fair Range LP(D, w) and a collection of super balls $\{\mathcal{P}(v)\}_{v \in D'}$ that satisfies,

1. *For every $v \in D', \mathcal{B}(v) \subseteq \mathcal{P}(v)$*
2. *For every $v \in D'$ and $u \in \mathcal{P}(v) \backslash \mathcal{B}(v)$, $x_{vu} > 0$ only if $\sum_{u \in \mathcal{B}(v)} y_u < 1$. Similarly, for every $v \in D'$ and $u \in F \backslash \mathcal{P}(v)$, $x_{vu} > 0$ only if $\sum_{u \in \mathcal{P}(v)} y_v < 1$*
3. *For every $v \in D'$, if $x_{vu} > 0$, then either $u \in \mathcal{P}(v)$ or $u \in \mathcal{B}(v')$ where v' denotes the nearest location to v in D'*
4. *For every v in D', $\sum_{u \in \mathcal{P}(v)} x_{vu} \geq \sum_{u \in \mathcal{B}(v)} x_{vu} \geq 1/2$*
5. *For every $v \in D'$, $u \in \mathcal{P}(v) \backslash \mathcal{B}(v)$, $d(u, v) \leq 2d(v, v')$*
6. *The set of super balls, $\{\mathcal{P}(v)\}_{v \in D'}$ are disjoint*

Lemma

1. *The optimal fractional solution of Structured LP*(D', w') *is a valid solution for fair range clustering on* (D', w') *and has cost at most* $e^{O(p)} \cdot OPT_D$.
2. *Consider a half-integral solution* $\tilde{y}$ *of Structured LP*(D', w') *of cost* z. *Then,* $\tilde{y}$ *is a feasible solution for Fair Range LP*(D', w') *with cost at most* $\left(\frac{3}{2}\right)^p \cdot z$.
3. *The matrix corresponding to the constraints of Structured LP is a TU matrix.*

If A is TU and b is integral, then for any cost vector c, the linear programs of the form $\{\min cx | Ax \geq b, x \geq 0\}$ has integral optima.

Rounding Algorithm

Algorithm 1 Partitioning Facilities.

Input: A set of locations D', half-integral vector y

for all location $v_i \in D'$ **do**

  $R_i \leftarrow$ the minimum assignment cost of a unit of demand at v_i w.r.t. y: i.e., $R_i = \frac{1}{2}(d(v_i, u_{i_1})^p + d(v_i, u_{i_1})^p)$ where u_{i_1}, u_{i_2} are respectively the primary and secondary facilities serving v_i

  $S_i \leftarrow \{u_{i_1}\} \cup \{u_{i_2}\}$

end for

$D'' \leftarrow D', \overline{D} \leftarrow \emptyset$

while D'' is nonempty **do**

  let $v_i \leftarrow \text{argmin}_{v_j \in D''} R_j$

  add v_i to $\overline{D}$

  remove all locations $v_j \in D''$ such that $S_j \cap S_i \neq \emptyset$

end while