Efficient Online Learning for Dynamic k-Clustering

Dimitris Fotakis, National Technical University of Athens
Georgios Piliouras Singapore University of Technology and Design
Stratis Skoulakis, Singapore University of Technology and Design

ICML 2021

Clustering problems
Open k centers minimizing the distance from n clients.

- n clients on a metric space $\mathrm{d} : \mathrm{V} \times \mathrm{V} \mapsto \mathbb{R}$.
- Open k centers, $\mathrm{F} \subseteq \mathrm{V}$ with $|F| = k$.
- **Each client connects to the closest center,** $\mathrm{d}(i, F) = \min_{j \in F} d_{ij}$,

$$\textbf{minimize} \quad \underbrace{\left(\sum_{i=1}^{n} \mathrm{d}^p(i, F) \right)^{1/p}}_{\text{the } p\text{−norm of distances}} \quad \text{where } p \geq 1$$

Extensively studied for various p-values.

- k-median for $p = 1$.
- k-means for $p = 2$.
- k-center for $p = \infty$.

k-Clustering Problems

Clustering problems
Open k centers minimizing the distance from n clients.

- n clients on a metric space $\mathrm{d} : \mathrm{V} \times \mathrm{V} \mapsto \mathbb{R}$.
- Open k centers, $\mathrm{F} \subseteq \mathrm{V}$ with $|F| = k$.
- **Each client connects to the closest center,** $\mathrm{d}(i, F) = \min_{j \in F} d_{ij}$,

$$\textbf{minimize} \quad \underbrace{\left(\sum_{i=1}^{n} \mathrm{d}^p(i, F)\right)^{1/p}}_{\text{the } p-\text{norm of clients' distances}} \quad \text{where } p \geq 1$$

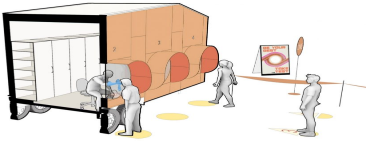
Many efficient approximation algorithms.

k-Clustering under Uncertainty

In most practical settings, the positions of the clients are unknown,

- clients are located according to some unknown probability distribution.
- clients change positions over time.

example
Locate the k Covid-Test Units over a city.



k-Clustering under Uncertainty

In most practical settings, the positions of the clients are unknown,

- clients are located according to some unknown probability distribution.
- clients change positions over time.

Place k-centers in such agnostic settings??

We can cast **k-Clustering** as Online Learning Problem.

k-Clustering as Online Learning Problem

A learner selects the positions of the centers and an adversary selects the positions of the clients.

At each round $t \geq 1$,

1. The learner chooses the positions of the centers, $F_t \subseteq V$ with $|F_t| = k$.
2. The adversary selects the positions of the clients, denoted as R_t.
3. The learner suffers the connection cost of the clients,

$$C(R_t, F_t) = \left(\sum_{i \in R_t} d^p(i, F_t) \right)^{1/p}$$

where $d(i, F_t) = \min_{j \in F_t} d_{ij}$.

Online Learning Algorithm
Takes as input the past positions of the clients $R_1, \ldots, R_{t-1}$ and outputs the positions of the centers at round t (denoted as F_t) .

Goal of Learner:
Select the centers according to an Online Learning Algorithm minimizing connection cost.

- If the learner uses the Multiplicative Weight Update algorithm then

$$\text{Overall Connection Cost} \sim \text{Optimal Cost!!}$$

Online k-Clustering

- If the learner uses the Multiplicative Weight Update algorithm then

$$\underbrace{\sum_{t=1}^{T} C(R_t, F_t)}_{\text{the overall connection cost}} \leq \underbrace{\min_{F^*} \sum_{t=1}^{T} C(R_t, F^*)}_{\text{the optimal connection cost}} + O(\sqrt{T})$$

- MWU needs $\Theta\left(|V|^k\right)$ time and space!! (Highly Inefficient)

Can we achieve similar performance guarantees in polynomial-time?

Our Results

- We design a polynomial-time online learning algorithm $\mathrm{poly}(|V|, k)$ such that

$$\sum_{t=1}^{T} C(R_t, F_t) \leq \rho \cdot \sum_{t=1}^{T} C(R_t, F^*) + O(\sqrt{T})$$

  for $\rho = \min(k, \#\text{clients per round})$.

- constant ρ (independent form $k, \#$clients per round) cannot be achieved with polynomial-time online learning algorithms.

Thanks You :)