Approximation-Preserving Reduction of k-Means Clustering with a Given Center to k-Means Clustering

Alexander Ageev

Sobolev Institute of Mathematics,
4 Koptyug Ave., 630090 Novosibirsk, Russia
ageev@math.nsc.ru

We consider the following clustering problem introduced by A. Kelmanov and studied in a series of papers by him and his coauthors (see [1] for the references).

Problem k-Means+Fixed Center.

Instance: A set X consisting of n points $\mathbb{R}^d$ and a positive integer k.

Goal: Find a family of mutually disjoint subsets $C_1, \dots C_k \subseteq X$ minimizing the function

$$\sum_{i=1}^{k} \sum_{x \in C_i} \|x - z(C_i)\|^2 + \sum_{x \in X \setminus (\bigcup_i C_i)} \|x\|^2$$

where $z(C_i)$ stands for the centroid of C_i.

Problem k-Means+Fixed Center is a natural modification of the classical k-Means problem:

Problem k-Means.

Instance: A set X consisting of n points $\mathbb{R}^d$ and a positive integer k.

Goal: Find a partition $C_1, \dots C_k$ of X minimizing the function

$$\Psi(\mathcal{C}) = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - z(C_i)\|^2$$

Theorem 1. *Let $\mathcal{A}$ be an algorithm solving k-means with approximation ratio $1 + \delta$ and time complexity $T(d, n, k, 1/\delta)$. Then problem k-means+fixed center can be solved with approximation ratio $(1 + \varepsilon)(1 + \delta)$ in time $T(d, n(1 + 1/\varepsilon), k + 1, 1/\delta)$.*

This theorem allows to carry known algorithmic results from k-Means problem to k-Means+Fixed Center problem.

Acknowledgments. This work was supported by the Russian Foundation for Basic Research (project 15-01-00462).

References

1. Alexander V. Kelmanov and Vladimir I. Khandeev. A Randomized Algorithm for Two-Cluster Partition of a Set of Vectors. Computational Mathematics and Mathematical Physics 55(2), 2015, 330–339.