Improved Local-Search Algorithm for k-Median

Vincent Cohen-Addad[1], Anupam Gupta[2], Hoon Oh[2],Lunjia Hu[3] David Saulpic[1]

[1] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
vcohenad@gmail.com, david.saulpic@lip6.fr

[2] Carnegie Mellon University, USA
anupamg@cs.cmu.edu, hooh@andrew.cmu.edu

[3] Stanford University (work done while visiting CMU)
lunjia.hu@stanford.edu

Mots-clés : *Approximation Algorithm, Clustering.*

1 Introduction

In this work, we propose a “non-oblivious” local search algorithm for the k-median problem and analyze its performance. In the k-median problem we are given a finite metric space $(\mathcal{X}, d)$, where $\mathcal{X} = \mathcal{C} \cup \mathcal{F}$, and the goal is to pick k medians $F \subseteq \mathcal{F}$ to minimize $\sum_{c \in \mathcal{C}} d(c, F)$. The current-best approximation algorithm for the problem is a $2.675+\epsilon$ factor due to Byrka et al. [3], improving on a breakthrough $2.732+\epsilon$ factor due to Li and Svensson [5]. These algorithms are based the so-called bi-point solutions given by the primal-dual framework.

Prior to these works, the best approximation factor had been $(3+\varepsilon)$ due to Arya et al. [2]. This result was based on a natural swap-based local search algorithm, which moved from the current solution F to a new solution F' only if the k-median cost of F' was strictly better than that of F, and moreover $|F \Delta F'| \leq O(1/\epsilon)$. The analysis shows that the *locality ratio*, i.e., the ratio of the cost of the worst local optimum with respect to the cost of the global optimum, is at most $(3+\epsilon)$. Arya et al. also showed instances with a matching locality gap for this algorithm.

2 Our Contribution

A natural question is : what if we perform the local search with respect to some other “surrogate” measure $\Phi(F)$ that allows us to move out of these bad local minima, yet is close enough to the original objective function so that local-optimality guarantees with respect to Φ can imply a good approximation for the original problem. Such a local-search procedure has been called a *non-oblivious* local search, and has been successful in several settings, e.g. [4, 1].

In this work we show that considering the potential function

$$\Phi(F) = \sum_{c \in \mathcal{C}} \Big(d(c, F) + \beta \min \left\{ \alpha\, d(c, F), \text{distance to second-closest facility in } F \right\} \Big),$$

for some α and β, allows us to reach a $2.82+\epsilon$ approximation factor. This improves the long-standing barrier for local-search algorithms, but do no improve the state-of-the-art.

The formal algorithm is therefore :

Algorithm 1 Local Search for finding k clusters

Input : A metric space and associated distance function $d(\cdot)$, an n-element set C of points, error parameter $\epsilon > 0$, positive integer parameter k
$S \leftarrow$ Arbitrary size-k set of points
while $\exists$ S' s.t. $|S'| \leq k$ **and** $|S \setminus S'| + |S' \setminus S| = O(1/\epsilon)$ **and** $\Phi(S') \leq \Phi(S)$ **do**
    $S \leftarrow S'$
end while
Output : S

Interestingly, the only locality-gap example we know has a ratio $3 - 2\beta$, which is 2.6 with our choice of β. Hence, our analysis may be far from being optimal, and there remains rooms for improvement.

3 Techniques

To analyse the local optima of the algorithm, we define some *swaps* involving facilities of the local solution L and the optimal one O. For instance, since L is a local optimum, for any $f \in L$ and $g \notin L$, it holds that

$$\Phi(L) - \phi(L - \{f\} + \{g\}) \leq 0$$

The proof of the approximation guarantee of the algorithm consists of a combination of such inequalities, carefully weighted and chosen. In order to define those swaps, we introduce the functions $\eta_1 : O \to L$ (resp. η_2) that associates a facility of O with its closest (resp. second closest) facility of L. For all facilities f of L, we consider the two swaps exchanging f against $\eta_1^{-1}(f)$ and $\eta_2^{-1}(f)$. Note that those swaps may be unbalanced (when $|\eta_1^{-1}(f)| > 1$) or too big $|\eta_1^{-1}(f)| > O(1/\epsilon)$: we handle the first case by adding some other facilities of L as needed, and deal separately with the second case.

As in Arya et al. **??**, we need to introduce swaps involving more than a single facility of L. To define those swaps, we introduce a random graph, where every facility of L is connected to its closest facility of O and every facility f^* of O is randomly connected to $\eta_1(f^*)$ or to $\eta_2(f^*)$. For every connected component of this graph, we create a swap exchanging all facilities of L of this component to the one of O.

Those types of swaps turn out to be enough to improve the approximation ratio of local search to $2.82 + \epsilon$.

Références

[1] Paola Alimonti. Non-oblivious local search for graph and hypergraph coloring problems. In *International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 167–180. Springer, 1995.

[2] Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristics for k-median and facility location problems. *SIAM Journal on computing*, 33(3) :544–562, 2004.

[3] Jarosław Byrka, Thomas Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for k-median, and positive correlation in budgeted optimization. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 737–756. SIAM, 2014.

[4] Yuval Filmus and Justin Ward. Monotone submodular maximization over a matroid via non-oblivious local search. *SIAM Journal on Computing*, 43(2) :514–542, 2014.

[5] Shi Li and Ola Svensson. Approximating k-median via pseudo-approximation. *SIAM Journal on Computing*, 45(2) :530–547, 2016.