Discrete Optimization 22 (2016) 312–327

Contents lists available at ScienceDirect

Discrete Optimization

www.elsevier.com/locate/disopt

An approximation algorithm for the Euclidean incremental median problem☆

Vladimir Shenmaier

Sobolev Institute of Mathematics, Novosibirsk, Russia

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form 25 March 2016
Accepted 25 August 2016
Available online 17 September 2016

Keywords:
Incremental medians
Euclidean medians
Hierarchical clustering
Approximation algorithm
Online algorithm

ABSTRACT

In the incremental version of the k-median problem, we find a sequence of facility sets $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n$, where each F_k contains at most k facilities. This sequence is said to be δ-*competitive* if the cost of each F_k is at most δ times the optimum cost of k facilities. The best deterministic (randomized) algorithm available for the metric space has a competitive ratio of 8 (7.656). The best one for the one-dimensional problem finds a 5.828-competitive sequence. We give a 7.076-competitive solution for the high-dimensional Euclidean space.

1. Introduction

In this paper, we study the incremental version of the well-known k-median problem. The feature of this version is that the total number k of facilities to be placed is unknown. Instead, we must place facilities one at a time so that at each step the facilities already placed constitute an optimal or approximate solution for the median problem of the corresponding cardinality.

An instance of the problem is specified by a finite set of customers C, a finite set of facilities F, the distance $dist(u, f) \geq 0$ defined for each customer u and each facility f, and the weight $w(u) \geq 0$ defined for each customer u. The *cost* of an arbitrary set of facilities X is the weighted sum

$$cost(X) = \sum_{u \in C} w(u)\, dist(u, X),$$

where the distance $dist(u, X)$ from the customer u to the set X is defined as the distance from u to the nearest facility in X.

☆ This research is supported by RFBR (projects 13-07-00070, 15-01-00462 and 15-01-00976).
E-mail address: shenmaier@mail.ru.

http://dx.doi.org/10.1016/j.disopt.2016.08.005

In the classical k-median problem, we are given a positive integer k, the objective is to find a k-element facility set of the minimum cost. In the incremental version of the problem, we find an incremental sequence of facility sets $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n$, where each F_k contains at most k facilities, n is the cardinality of F. This sequence is called *δ-competitive* if each F_k is a δ-approximate k-median:

$$cost(F_k) \leq \delta\, cost(X)$$

for any k-element facility set X. We say that an algorithm for the incremental problem is δ-competitive, or has a *competitive ratio* of δ, if it produces a δ-competitive solution $F_1, F_2, \ldots, F_n$ for any instance of the problem.

Related work

The first constant-factor solution for the metric incremental median problem was proposed by Mettu and Plaxton [1]. They obtained a linear-time algorithm with competitive ratio 29.86. Chrobak et al. [2] (and, independently, Lin et al. [3]) proved that each instance has an 8-competitive incremental medians sequence. This produces an algorithm with competitive ratio $8c$, where c is the approximation ratio of available k-median problem solutions. Chrobak and Hurand [4] proved, via a probabilistic argument, the existence of an improved solution with competitive ratio at most 7.656 which produces a randomized algorithm with competitive ratio $7.656c$. They also showed that in the general metric case, no competitive ratio better than 2.01 is possible. Note that, for $c = 1$, the above algorithms become non-polynomial since they use optimal k-medians. Lin et al. [3] gave a polynomial-time algorithm with competitive ratio 16. For the one-dimensional case, a 5.828-competitive solution was proposed in the work [5].

Our result

We present a $7.076c$-competitive algorithm for the Euclidean problem. In this case, both customers and facilities lie in the space $\mathbb{R}^d$ and the distance function is induced by Euclidean metric:

$$dist(u, f) = \|u - f\|_2 = \sqrt{\sum_{i=1}^{d}(u(i) - f(i))^2}.$$

If the dimension of space is fixed, we can use PTAS for Euclidean k-medians [6]. In this case, since $c = 1+\varepsilon$, our algorithm gives a $(7.076 + \varepsilon)$-competitive solution for the incremental median problem in polynomial time.

2. Algorithm description

The basic idea of the algorithm is similar to that of the algorithms from [2,5]. We firstly find c-approximate solutions $F_1^*, F_2^*, \ldots, F_n^*$ of the usual k-median problem for $k = 1, 2, \ldots, n$. Next, the algorithm determines the inextensible sequence of indices $i_1, i_2, \ldots, i_t$ in which $i_1 = 1$ and, for $k = 2, \ldots, t$, the index i_k is minimum such that

$$cost(F_{i_k}^*) < cost(F_{i_{k-1}}^*)/\alpha,$$

where α is a parameter of the algorithm, $\alpha = 2.04266$. Note that, when the previous algorithms are used, the analogous constant is 2 for the 8-competitive algorithm of [2] and $1 + \sqrt{2}$ for the one-dimensional algorithm of [5].

Then, we construct a partial solution for the incremental median problem corresponding to the indices i_k. To wit, we find an increasing sequence of sets $F_{i_1} \subseteq F_{i_2} \subseteq \cdots \subseteq F_{i_t}$ constructed recursively in the