Theoretical Computer Science 602 (2015) 149–157

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

A deterministic sublinear-time nonadaptive algorithm for metric 1-median selection

Ching-Lueh Chang [a,b,*]

[a] *Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan*
[b] *Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, Taiwan*

ARTICLE INFO

Article history:
Received 24 February 2015
Received in revised form 30 July 2015
Accepted 31 July 2015
Available online 5 August 2015
Communicated by S. Sen

Keywords:
1-Median selection
Sublinear-time algorithm
Metric space
k-Median selection
Closeness centrality

ABSTRACT

We give a deterministic $O(hn^{1+1/h})$-time $(2h)$-approximation nonadaptive algorithm for 1-median selection in n-point metric spaces, where $h \in \mathbb{Z}^+ \setminus \{1\}$ is arbitrary. Our proof generalizes that of Chang [2].

1. Introduction

A metric space (M, d) is a nonempty set M endowed with a function $d: M \times M \to [0, \infty)$ such that for all $x, y, z \in M$,

- $d(x, y) = 0$ if and only if $x = y$,
- $d(x, y) = d(y, x)$, and
- $d(x, y) + d(y, z) \geq d(x, z)$ (triangle inequality).

The METRIC 1-MEDIAN problem asks for a point in an n-point metric space (M, d) with the minimum average distance to other points. For $c \geq 1$, a point $\hat{p} \in M$ is said to be c-approximate for METRIC 1-MEDIAN if

$$\sum_{x \in M} d(\hat{p}, x) \leq c \cdot \min_{p \in M} \sum_{x \in M} d(p, x).$$

An algorithm for METRIC 1-MEDIAN is nonadaptive if the sequence of distances that it inspects depends only on M but not on d. Because there are $n(n-1)/2$ nonzero distances, "sublinear-time" will mean "$o(n^2)$-time."

Indyk [5,6] shows that METRIC 1-MEDIAN has a Monte-Carlo $O(n/\epsilon^2)$-time $(1+\epsilon)$-approximation algorithm for each $\epsilon > 0$. In $\mathbb{R}^D$, where $D \geq 1$, METRIC 1-MEDIAN has a Monte-Carlo $O(2^{\text{poly}(1/\epsilon)} D)$-time $(1+\epsilon)$-approximation algorithm for all $\epsilon > 0$ [7]. Many other algorithms are known for k-median selection [1,4,7]. For example, Guha et al. [4] give a deterministic,

* Correspondence to: Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan.
E-mail address: clchang@saturn.yzu.edu.tw.

http://dx.doi.org/10.1016/j.tcs.2015.07.058
0304-3975/

$O(n^{1+\epsilon})$-time, $O(n^{\epsilon})$-space, $2^{O(1/\epsilon)}$-approximation and one-pass algorithm as well as a Monte-Carlo algorithm for k-median selection in metric spaces, where $\epsilon > 0$.

We show that METRIC 1-MEDIAN has a deterministic $O(hn^{1+1/h})$-time $(2h)$-approximation nonadaptive algorithm for all $h \in \mathbb{Z}^+ \setminus \{1\}$, generalizing the following theorems:

Theorem 1. *([2])* METRIC 1-MEDIAN *has a deterministic $O(n^{1.5})$-time 4-approximation nonadaptive algorithm.*

Theorem 2. *([8]) For each $h \in \mathbb{Z}^+ \setminus \{1\}$,* METRIC 1-MEDIAN *has a deterministic $O(hn^{1+1/h})$-time $(2h)$-approximation (adaptive) algorithm.*[1]

When n is a perfect square and $h = 2$, our proof is equivalent to that of Theorem 1 [2]. Chang [3] shows that METRIC 1-MEDIAN has no deterministic $o(n^2)$-query $(4 - \Omega(1))$-approximation algorithms (where an algorithm's query complexity is the number of distances that it inspects). So the approximation ratio of 4 in Theorem 1 cannot be improved to a smaller constant.

2. Main result

Let $(\{0, 1, \ldots, n-1\}, d)$ be a metric space and $h \in \mathbb{Z}^+ \setminus \{1\}$. Our goal is to give a deterministic $O(hn^{1+1/h})$-time $(2h)$-approximation nonadaptive algorithm for METRIC 1-MEDIAN. To this end, we will design a function $\tilde{d}\colon \{0, 1, \ldots, n-1\}^2 \to [0, \infty)$ with the following properties:

(1) A 1-median with respect to $\tilde{d}$ is a $(2h)$-approximate 1-median with respect to d.
(2) There exists a computable set $S(n, h) \subseteq \{0, 1, \ldots, n-1\}^2$ (depending only on n and h) of size $O(hn^{1+1/h})$ such that $\tilde{d}(\cdot, \cdot)$ can be computed deterministically given $d(i, j)$ for $(i, j) \in S(n, h)$.

Items (1)–(2) imply that simply outputting a 1-median with respect to $\tilde{d}$ yields a deterministic $O(hn^{1+1/h})$-query $(2h)$-approximation nonadaptive algorithm.

A rough and informal intuition for constructing $\tilde{d}$ is in order. For independent and uniformly random points $\boldsymbol{u}$ and $\boldsymbol{v}$, we will let

$$\tilde{d}(\boldsymbol{u}, \boldsymbol{v}) = \sum_{k=0}^{h-1} d\left(c_k, c_{k+1}\right)$$

for a suitable sequence of uniformly random (but not independent) points, $c_0 = \boldsymbol{u}, c_1, c_2, \ldots, c_{h-1}, c_h = \boldsymbol{v}$. This and the triangle inequality for d imply that for any 1-median OPT,

$$\tilde{d}(\boldsymbol{u}, \boldsymbol{v}) \le \sum_{k=0}^{h-1} \left(d(\mathrm{OPT}, c_k) + d\left(\mathrm{OPT}, c_{k+1}\right)\right),$$

whose right-hand side is a sum of $2h$ distances from OPT to uniformly random points. This makes item (1) plausible. To make item (2) plausible as well, we will require (c_k, c_{k+1}), where $k \in \{0, 1, \ldots, h-1\}$, to take a certain form that is taken by only $O(hn^{1+1/h})$ pairs of points.

We now proceed with formal constructions. Let $t \overset{\text{def.}}{=} \lceil n^{1/h} \rceil$. For all $j \in \{0, 1, \ldots, n-1\}$, denote the (unique) t-ary representation of j by

$$\left(s_{h-1}(j), s_{h-2}(j), \ldots, s_0(j)\right) \in \{0, 1, \ldots, t-1\}^h,$$

i.e.,

$$\sum_{\ell=0}^{h-1} s_\ell(j) \cdot t^\ell = j. \tag{1}$$

For all $i, j \in \{0, 1, \ldots, n-1\}$,

$$\tilde{d}(i, i + j \bmod n) \overset{\text{def.}}{=} \sum_{k=0}^{h-1} d\left(i + \sum_{\ell=h-k}^{h-1} s_\ell(j) \cdot t^\ell \bmod n,\, i + \sum_{\ell=h-1-k}^{h-1} s_\ell(j) \cdot t^\ell \bmod n\right). \tag{2}$$

By convention, empty sums vanish, e.g., $\sum_{\ell=h}^{h-1} s_\ell(j) \cdot t^\ell = 0$.

[1] The time complexity of $O(hn^{1+1/h})$ is originally presented as $O(n^{1+1/h})$ because h is independent of n. We include the $O(h)$ factor, which is implicit in the original proof, for ease of comparison.