Theoretical Computer Science 540–541 (2014) 27–39

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Online Sum-Radii Clustering[☆]

Dimitris Fotakis [a,*], Paraschos Koutris [b]

[a] School of Electrical and Computer Engineering, National Technical University of Athens, 157 80 Athens, Greece

[b] Computer Science and Engineering, University of Washington, USA

ARTICLE INFO

Keywords:
Online algorithms
Competitive analysis
Sum-k-Radii Clustering

ABSTRACT

In Online Sum-Radii Clustering, n demand points arrive online and must be irrevocably assigned to a cluster upon arrival. The cost of each cluster is the sum of a fixed opening cost and its radius, and the objective is to minimize the total cost of the clusters opened by the algorithm. We show that the deterministic competitive ratio of Online Sum-Radii Clustering for general metric spaces is $\Theta(\log n)$, where the upper bound follows from a primal–dual algorithm and holds for general metric spaces, and the lower bound is valid for ternary Hierarchically Well-Separated Trees (HSTs) and for the Euclidean plane. Combined with the results of (Csirik et al., MFCS 2010), this result demonstrates that the deterministic competitive ratio of Online Sum-Radii Clustering changes abruptly, from constant to logarithmic, when we move from the line to the plane. We also show that Online Sum-Radii Clustering in metric spaces induced by HSTs is closely related to the Parking Permit problem introduced by (Meyerson, FOCS 2005). Exploiting the relation to Parking Permit, we obtain a lower bound of $\Omega(\log \log n)$ on the randomized competitive ratio of Online Sum-Radii Clustering in tree metrics. Moreover, we present a simple randomized $O(\log n)$-competitive algorithm and a deterministic $O(\log \log n)$-competitive algorithm for the fractional version of the problem.

1. Introduction

In clustering problems, we seek a partitioning of n demand points into k groups, or *clusters*, so that a given objective function, that depends on the distance between points in the same cluster, is minimized. Typical examples are the k-Center problem, where we minimize the maximum cluster diameter, the Sum-k-Radii problem, where we minimize the sum of cluster radii, and the k-Median problem, where we minimize the total distance of points to the nearest cluster center. These are fundamental problems in Computer Science, with many important applications, and have been extensively studied from an algorithmic viewpoint (see e.g., [19] and the references therein).

In this work, we study an online clustering problem closely related to Sum-k-Radii. In the online setting, the demand points arrive one-by-one and must be irrevocably assigned to a cluster upon arrival. Specifically, when a new demand point u arrives, if it is not covered by an open cluster, the algorithm has to open a new cluster covering u and to assign u to it. Opening a new cluster means that the algorithm must irrevocably fix the center and the radius of the new cluster. We

☆ This work was supported by the project AlgoNow, co-financed by the European Union (European Social Fund — ESF) and Greek national funds, through the Operational Program "Education and Lifelong Learning", under the research funding program THALES. Part of this work was done while P. Koutris was with the School of Electrical and Computer Engineering, National Technical University of Athens, Greece. An extended abstract of this work appeared in the Proceedings of the 37th Symposium on Mathematical Foundations of Computer Science (MFCS 2012), Branislav Rovan, Vladimiro Sassone, and Peter Widmayer (Editors), Lecture Notes in Computer Science 7464, pp. 395–406, Springer, 2012.

* Corresponding author. Tel.: +30 210 7724302.
E-mail addresses: fotakis@cs.ntua.gr (D. Fotakis), pkoutris@cs.washington.edu (P. Koutris).

0304-3975/$ – see front matter
http://dx.doi.org/10.1016/j.tcs.2013.03.010

emphasize that once formed, clusters cannot be merged, split, or have their center or radius changed. The goal is to open few clusters with a small sum of radii. However, instead of requiring that at most k clusters open, which would lead to an unbounded competitive ratio, we follow [7,8] and consider a Facility-Location-like relaxation of Sum-k-Radii, called *Sum-Radii Clustering*. In Sum-Radii Clustering, the cost of each cluster is the sum of a fixed opening cost and its radius, and we seek to minimize the total cost of the clusters opened by the algorithm.

In addition to clustering and data analysis, Sum-Radii Clustering has applications to the problem of base station placement for the design of wireless networks where users are scattered to various locations (see e.g., [8,3,16]). In such problems, we place some wireless base stations and set up their communication range so that the communication demands are satisfied and the total setup and operational cost is minimized. A standard assumption is that the setup cost is proportional to the number of stations installed, and the operational cost for each station is proportional to the energy consumption, which is typically modeled by a low-degree polynomial of its range. In Sum-Radii Clustering, we study the particular variant where the operational cost has a linear dependence on the range.

1.1. Previous work

In the offline setting, Sum-k-Radii and the closely related problem of Sum-k-Diameters[1] have been thoroughly studied. Sum-k-Radii is **NP**-hard even in metric spaces of constant doubling dimension [15]. Gibson et al. [14] proved that Sum-k-Radii in Euclidean spaces of constant dimension is polynomially solvable and presented an $O(n^{\log \Delta \log n})$-time algorithm for Sum-$k$-Radii in general metric spaces, where Δ is the diameter [15]. As for approximation algorithms, Doddi et al. [10] proved that it is **NP**-hard to approximate Sum-k-Diameters in general metric spaces within a factor less than 2 and gave a bicriteria algorithm that achieves a logarithmic approximation using $O(k)$ clusters. Subsequently, Charikar and Panigraphy [7] presented a primal–dual $(3.504+\varepsilon)$-approximation algorithm for Sum-k-Radii in general metric spaces, which uses as a building block a primal–dual 3-approximation algorithm for Sum-Radii Clustering. Biló et al. [3] considered a generalization of Sum-k-Radii, where the cost is the sum of the α-th power of the clusters radii, for $\alpha \geq 1$, and presented a polynomial-time approximation scheme for Euclidean spaces of constant dimension.

Charikar and Panigraphy [7] also considered the incremental version of Sum-k-Radii. Similarly to the online setting, an incremental algorithm processes the demands one-by-one and assigns them to a cluster upon arrival. However, an incremental algorithm can also merge any of its clusters at any time. They presented an $O(1)$-competitive incremental algorithm for Sum-k-Radii that uses $O(k)$ clusters.

In the online setting, where cluster reconfiguration is not allowed, the Unit Covering and the Unit Clustering problems have received considerable attention. In both problems, the demand points arrive one-by-one and must be irrevocably assigned to unit-radius balls upon arrival so that the number of balls used is minimized. The difference is that in Unit Covering, the center of each ball is fixed when the ball is first used, while in Unit Clustering, there is no fixed center and a ball may shift and cover more demands. Charikar et al. [6] proved an upper bound of $O(2^d d \log d)$ and a lower bound of $\Omega(\log d/ \log\log\log d)$ on the deterministic competitive ratio of Unit Covering in d dimensions. The results of [6] imply a competitive ratio of 2 and 4 for Unit Covering on the line and the plane, respectively. The Unit Clustering problem was introduced by Chan and Zarrabi-Zadeh [5]. The deterministic competitive ratio of Unit Clustering on the line is at most $5/3$ [11] and no less than $8/5$ [12]. Unit Clustering has also been studied in d-dimensions with respect to the L_∞ norm, where the competitive ratio is at most $\frac{5}{6}2^d$, for any d, and no less than $13/6$, for $d \geq 2$ [11].

Departing from this line of work, Csirik et al. [8] studied online Clustering to minimize the sum of the Setup Costs and the Diameters of the clusters, or CSDF in short. Motivated by the difference between Unit Covering and Unit Clustering, they considered three models, the strict, the intermediate, and the flexible one, depending on whether the center and the radius of a new cluster are fixed at its opening time. Csirik et al. only studied CSDF on the line and proved that its deterministic competitive ratio is $1+\sqrt{2}$ for the strict and the intermediate model and $(1+\sqrt{5})/2$ for the flexible model. Recently, Divéki and Imreh [9] studied online clustering in two dimensions to minimize the sum of the setup costs and the area of the clusters. They proved that the competitive ratio of this problem lies in $(2.22, 9]$ for the strict model and in $(1.56, 7]$ for the flexible model.

1.2. Contribution

Following [8], it is natural and interesting to study the online clustering problem of CSDF in metric spaces more general than the line, such as trees, the Euclidean plane, and general metric spaces. In this work, we consider the closely related problem of Online Sum-Radii Clustering (OnlSumRad) and give upper and lower bounds on its deterministic and randomized competitive ratio for general metric spaces and for the Euclidean plane. We restrict our attention to the strict model of [8], where the center and the radius of each new cluster are fixed at opening time. To justify our choice, we show that a c-competitive algorithm for the strict model implies an $O(c)$-competitive algorithm for the intermediate and the flexible model.

[1] These problems are closely related in the sense that a c-competitive algorithm for Sum-k-Radii implies a $2c$-competitive algorithm for Sum-k-Diameters, and vice versa.