Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Approximating snowflake metrics by trees

William Leeb [a,b]

[a] Dept. of Mathematics, Yale University, New Haven, CT 06511, United States [1]
[b] PACM, Princeton University, Princeton, NJ 08554, United States

ARTICLE INFO

Article history:
Received 25 January 2015
Received in revised form 4 November 2015
Accepted 7 October 2016
Available online xxxx
Communicated by Charles K. Chui

MSC:
68W05
68W25
05C05
54E35

Keywords:
Tree metric
Partition trees
Tree approximation
Snowflake metric
Dimension
Spaces of homogeneous type
Earth Mover's Distance
EMD

ABSTRACT

Tree metrics are encountered throughout pure and applied mathematics. Their simple structure makes them a convenient choice of metric in many applications from machine learning and computer science. At the same time, there is an elegant theory of harmonic analysis with respect to tree metrics that parallels the classical theory.

A basic question in this field, which is of both theoretical and practical interest, is how to design efficient algorithms for building trees with good metric properties. In particular, given a finite metric space, we seek a random family of dominating tree metrics approximating the underlying metric in expectation. For general metrics, this problem has been solved: on the one hand, there are finite metric spaces that cannot be approximated by trees without incurring a distortion logarithmic in the size of the space, while the tree construction of Fakcharoenphol, Rao, and Talwar (FRT, 2003) shows how to achieve such a logarithmic error for arbitrary metrics.

Since a distortion that grows even logarithmically with the size of the set may be too large for practical use in many settings, one naturally asks if there is a more restricted class of metrics where one can do better. The main result of this paper is that certain random family of trees already studied in the computer science literature, including the FRT trees, can be used to approximate snowflake metrics (metrics raised to a power less than 1) with expected distortion bounded by its doubling dimension and the degree of snowflaking. We also show that without snowflaking, the metric distortion can be bounded by a term logarithmic in the distance being approximated and linear in the dimension.

We also present an optimal algorithm for building a single FRT tree, whose running time is bounded independently of all problem parameters other than the number of points. We conclude by demonstrating our theoretical results on a numerical example, and applying them to the approximation of the Earth Mover's Distance between probability distributions.

E-mail address: wleeb@math.princeton.edu.

[1] Previous address.

http://dx.doi.org/10.1016/j.acha.2016.10.002
1063-5203/© 2016 Published by Elsevier Inc.

1. Introduction

Tree metrics are an especially simple kind of distance function that appear throughout pure and applied mathematics. Informally, tree metrics are derived by breaking the metric space into a tree $\mathcal{T}$ of nested subsets F, called folders, and assigning each folder a diameter $w(F)$. The distance $d_{\mathcal{T}}(x, y)$ between any two points x and y is then the diameter of the smallest folder containing both points.

There is an extensive theory of harmonic analysis for tree metrics that parallels the classical Euclidean theory. This theory allows us to adapt signal-processing type algorithms to data sets of much more varied structure, and has proven useful in a wide array of problems in machine learning [1–3]. Tree metrics' simple structure also yields fast algorithms for metric tasks from computer science, such as nearest neighbor searches, the k-server problem, distributed paging, the vehicle routing problem, and many more [4,5].

Unfortunately, it is rarely the case that the "natural" metric for a given problem in machine learning or computer science will be a tree metric. A basic goal in metric space theory, therefore, is to approximate arbitrary finite metrics by tree metrics. Of course, the extreme simplicity of tree metrics makes it implausible that an arbitrary metric could be well-approximated by a single tree metric. We therefore consider a modified problem, namely finding a probability distribution over tree metrics so that the expected tree distance yields a good approximation, and such that it is computationally feasible to draw a tree from the distribution.

The formal problem, as considered in [6,4,5,7–9] and elsewhere is as follows. Given a finite metric space (X, d), we seek a family of trees $\mathcal{T}$ and corresponding tree metrics $d_{\mathcal{T}}$ that have the following properties:

1. Each tree metric is *dominating*; that is,
$$d(x, y) \le d_{\mathcal{T}}(x, y) \tag{1}$$
for every $\mathcal{T}$ and for all $x, y \in X$.
2. The expected tree distance satisfies
$$\mathbb{E}_{\mathcal{T}}[d_{\mathcal{T}}(x, y)] \le K d(x, y) \tag{2}$$
for some constant $K \ge 1$.

Bartal's paper [4] describes such an explicit distribution over trees, where the constant K is of size $O(\log^2 n)$ where n denotes the number of points in X; this result was later improved to $K = O(\log n \log\log n)$ in [5]. With access to such a distribution over trees, many tasks that depend on the original metric can be performed with randomly drawn tree metrics instead, and then combined to produce an approximation to that task for the original metric. Bartal [4,5] discusses a number of such problems from computer science, while Charikar's paper [10] shows how this method can produce an approximation to the *Earth Mover's Distance*, a powerful metric between probability distributions widely used in machine learning [11–14]. We will go into more detail on this particular application in Section 5.

The question that naturally arises is: how small (that is, how close to 1) can we make the constant K from (2)? The paper of Fakcharoenphol, Rao, and Talwar [9] describes a randomized construction of partition trees whose constant of distortion K is of size $O(\log n)$. As there are metric spaces for which no family of trees can achieve a distortion smaller than $\Omega(\log n)$ [4], this result is optimal in the general case.

If n is large, however, a size $O(\log n)$ distortion can be too big for practical applications. Indeed, in a statistical or machine learning environment, if X is a data set drawn from a population about which we wish to make inferences, it is critical to be able to handle very large values of n, as well-designed statistical procedures perform better with increasing sample size.

In this paper we show that a broad class of metrics can in fact be approximated by trees with constant of distortion bounded independently of n. These metrics, known as *snowflake metrics*, are of the form $d(x, y)^{\alpha}$