Operations Research Letters 44 (2016) 370–373

Contents lists available at ScienceDirect

Operations Research Letters

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

A note on computing the center of uncertain data on the real line

Haitao Wang, Jingru Zhang *

Department of Computer Science, Utah State University, Logan, UT 84322, USA

ARTICLE INFO

Article history:
Received 21 August 2015
Received in revised form
14 March 2016
Accepted 14 March 2016
Available online 23 March 2016

Keywords:
One-center
Uncertain data
Histogram
Lowest point
Upper envelope
Algorithms

ABSTRACT

We consider the one-dimensional one-center problem on uncertain data. We are given a set $\mathcal{P}$ of n (weighted) uncertain points on a real line L and each uncertain point is specified by a probability density function that is a piecewise-uniform function (i.e., a histogram). The goal is to find a point c (the *center*) on L such that the maximum expected distance from c to all uncertain points of $\mathcal{P}$ is minimized. We present a linear-time algorithm for this problem.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In the real world, data is often associated with uncertainty due to their essence, the measurement inaccuracy, sampling discrepancy, resource limitation, etc. [6,9]. In this paper, we consider the one-dimensional one-center problem on uncertain data, defined as follows.

Let L be a real line. Without loss of generality, we assume L is the x-axis. Let $\mathcal{P}$ be a set of n uncertain points $\{P_1, P_2, \ldots, P_n\}$ on L, where each uncertain point $P_i \in \mathcal{P}$ is specified by its probability density function (pdf) $f_i: \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$, which is a piecewise-uniform function (i.e., a histogram), consisting of at most $m+1$ pieces (e.g., see Fig. 1). More specifically, for each P_i, there are m sorted x-coordinates $x_{i1} < x_{i2} < \cdots < x_{im}$ and $m-1$ nonnegative values $y_{i1}, y_{i2}, \ldots, y_{i,m-1}$ such that $f_i(x) = y_{ij}$ for $x_{ij} \le x < x_{i,j+1}$ with $1 \le j \le m-1$. For convenience of discussion, we assume $x_{i0} = -\infty$, $x_{i,m+1} = +\infty$, $y_{i0} = y_{im} = 0$, and $f_i(x) = 0$ for $x \in (-\infty, x_{i1}) \cup [x_{im}, +\infty)$.

As discussed in [1], such a histogram function f_i can be used to approximate any pdf with arbitrary precision. In particular, for the *discrete* case where each uncertain point has a finite number of discrete locations, each with a probability, it can also be incorporated by our histogram model using infinitesimal pieces at these locations. Thus, the discrete case is a special case of our histogram model.

With a little abuse of notation, for any (deterministic) point q on L, we also use q to denote the x-coordinate of q. For any uncertain point $P_i \in \mathcal{P}$, the *expected distance* from q to P_i is

$$\mathrm{Ed}(q, P_i) = \int_{-\infty}^{+\infty} f_i(x)|x - q|dx.$$

The goal of our *one-center* problem on $\mathcal{P}$ is to find a (deterministic) point c^* on L such that the maximum expected distance from c^* to all uncertain points of $\mathcal{P}$ is minimized, and c^* is called a *center* of $\mathcal{P}$.

The algorithm proposed in our previous work [18] can solve the problem in $O(mn \log mn + n \log n \log mn)$ time. In this paper, we present an $O(mn)$ time algorithm. Since the input size of the problem is $\Theta(mn)$, our algorithm runs in linear time, which is optimal.

We should point out that our algorithm is applicable to the *weighted case* of this problem where each uncertain point $P_i \in \mathcal{P}$ has a nonnegative multiplicative weight w_i and the weighted expected distance is considered (i.e., $\mathrm{Ed}(q, P_i) = w_i \cdot \int_{-\infty}^{+\infty} f_i(x)|x - q|dx$). To solve the weighted case, we can first reduce it to the above unweighted case by changing each value y_{ij} to $w_i \cdot y_{ij}$ for every $1 \le i \le n$ and $1 \le j \le m-1$, and then apply our algorithm for the unweighted case. The running time is still linear. Hence, we will focus our discussion on the unweighted case.

* Corresponding author.
E-mail addresses: haitao.wang@usu.edu (H. Wang), jingruzhang@aggiemail.usu.edu (J. Zhang).

http://dx.doi.org/10.1016/j.orl.2016.03.009
0167-6377/© 2016 Elsevier B.V. All rights reserved.

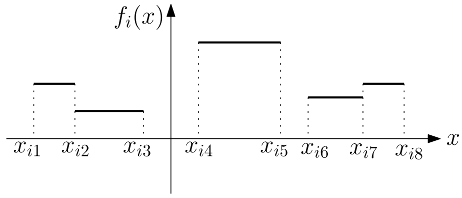

Fig. 1. Illustrating the pdf f_i of an uncertain point P_i with $m = 8$.

1.1. Related work

In our previous work [18], the problem of finding k centers for a set $\mathcal{P}$ of uncertain points on L was studied, and an algorithm of $O(mn \log mn + n \log k \log n \log mn)$ time was proposed. Therefore, when $k = 1$, the algorithm runs in $O(mn \log mn + n \log n \log mn)$ time as mentioned above. In addition, the discrete case of the above k-center problem (where each uncertain point P_i has m discrete locations, each with a probability) was solved in a faster way in $O(mn \log mn + n \log k \log mn)$ time [18], and the discrete one-center problem was solved in $O(mn)$ time [18]. Therefore, our new result in this paper for the one-center problem under the more general histogram model matches the previous result for the discrete case. We also studied the discrete one-center problem for uncertain points on tree networks and proposed a linear-time algorithm in [17].

The deterministic k-center problems are classical facility location problems that have been extensively studied. The problem is NP-hard in the plane [13]. Efficient algorithms were known for special cases, e.g., finding the smallest enclosing circle (i.e., the case $k = 1$) [12], k-center on trees [4,11,14]. The deterministic k-center problem in the one-dimensional space is solvable in $O(n \log n)$ time [5,7,14]. As shown in [18], the deterministic one-center problem in the one-dimensional space can be solved in linear time.

The k-center problems on uncertain data in high-dimensional spaces have also been considered. For example, approximation algorithms were given in [8] for different problem models, e.g., the assigned model that is somewhat similar to our problem model and the unassigned model which was relatively easy because it can be reduced to the corresponding deterministic problem [8]. Foul [10] studied the problem of finding the center in the plane to minimize the maximum expected distance from the center to all uncertain points, where each uncertain point has a uniform distribution in a given rectangle. Other facility location problems on uncertain data under various models, e.g., the minmax regret [2,3,16], have also been studied (see [15] for a survey).

1.2. Our techniques

To solve our one-center problem, based on observations, we reduce it to the following geometric problem. Let $\mathcal{H}$ be a set of n unimodal functions in the plane (i.e., when x changes from $-\infty$ to $+\infty$, it first monotonically decreases and then increases) and each function consists of m pieces with each piece being a parabolic arc. We wish to find the lowest point v^* in the upper envelope of the functions of $\mathcal{H}$. In the discrete version of the one-center problem, as shown in [18], each parabolic arc of every function of $\mathcal{H}$ is simply a line segment, and thus, v^* can be found by applying Megiddo's linear-time linear programming algorithm [12]. In our problem, however, since the parabolic arcs of the functions of $\mathcal{H}$ may not be line segments, Megiddo's algorithm in [12] does not work any more. We present a new prune-and-search technique that can compute v^* in $O(mn)$ time. This result immediately leads to a linear time algorithm for our one-center problem on $\mathcal{P}$.

Comparing with the linear programming problem, the above geometric problem is more general (i.e., the linear programming problem is a special case of our problem). Our linear time algorithm for the problem may be interesting in its own right and may find other applications as well. In fact, our result can be extended to more general unimodal functions (see the discussions in Section 4).

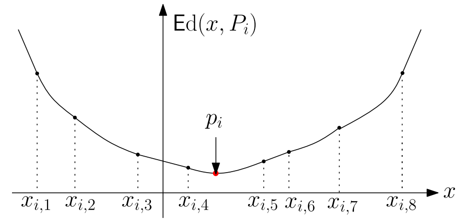

Fig. 2. Illustrating the expected distance function $\mathsf{Ed}(x, P_i)$ for an uncertain point P_i with $m = 8$; $\mathsf{Ed}(x, P_i)$ is monotonically decreasing for $x \in (-\infty, p_i]$, and increasing for $x \in [p_i, -\infty)$.

2. Preliminaries

A function $g : \mathbb{R} \to \mathbb{R}$ is a *unimodal* if there exists a value x' such that for any $x_1 < x_2$, $g(x_1) \geq g(x_2)$ holds if $x_2 \leq x'$ and $g(x_1) \leq g(x_2)$ holds if $x' \leq x_1$, i.e., $g(x)$ is monotonically decreasing on $x \in (-\infty, x']$ and increasing on $x \in [x', +\infty)$.

Recall that L is the x-axis, and for any point q on L, we also use q to denote the x-coordinate of q. Therefore, the values of $\mathbb{R}$ correspond to the points of L. In the following, we will use "the values of $\mathbb{R}$" and "the points of L" interchangeably.

Consider any uncertain point P_i of $\mathcal{P}$. If we consider the expected distance $\mathsf{Ed}(x, P_i)$ as a function of the points x on L (or $x \in \mathbb{R}$), the following observation has been proved in [18]. Note that the m coordinates $x_{i1}, \ldots, x_{im}$ of P_i are already given sorted.

Lemma 1 *([18]). The function* $\mathsf{Ed}(x, P_i)$ *for* $x \in \mathbb{R}$ *is unimodal. More specifically, there exists a point* $p_i \in L$ *such that* $\mathsf{Ed}(x, P_i)$ *is monotonically decreasing on* $x \in (-\infty, p_i]$ *and increasing on* $x \in [p_i, +\infty)$ *(e.g., see Fig. 2). In addition,* $\mathsf{Ed}(x, P_i)$ *is a parabolic arc (of constant complexity) on the interval* $[x_{ik}, x_{i,k+1})$ *for each* $0 \leq k \leq m$*, and can be explicitly computed in* $O(m)$ *time.*

The point p_i in Lemma 1 is referred to as a *centroid* of P_i and can be easily computed in $O(m)$ time after $\mathsf{Ed}(x, P_i)$ is explicitly computed in $O(m)$ time [18]. In fact, a point $q \in L$ is a centroid of P_i if and only if $\int_{-\infty}^{q} f_i(x) = 0.5$ and $\int_{q}^{+\infty} f_i(x) = 0.5$. Note that the centroid of P_i may not be unique. This case happens when there exists an interval on the x-axis such that for any point q in this interval both $\int_{-\infty}^{q} f_i(x) = 0.5$ and $\int_{q}^{+\infty} f_i(x) = 0.5$ hold, and thus $\mathsf{Ed}(x, P_i)$ is a constant when x is in this interval and any point q in this interval is a centroid. If the centroid of P_i is not unique, then we use p_i to refer to an arbitrary such centroid. For a similar reason, the center of $\mathcal{P}$ may also not be unique, in which case our algorithm will find one such center.

The following corollary can be easily obtained based on Lemma 1 and binary search on the sorted list of $x_{i1}, x_{i2}, \ldots, x_{im}$.

Corollary 1 *([18]). For each uncertain point* P_i*, with* $O(m)$ *time preprocessing, the value* $\mathsf{Ed}(x', P_i)$ *for any query value* x' *can be computed in* $O(\log m)$ *time.*

Let $\mathcal{H}$ denote the set of all functions $\mathsf{Ed}(x, P_i)$ for $i = 1, 2, \ldots, n$. Since the center c^* of $\mathcal{P}$ is a point on L that minimizes the value $\max_{i=1}^{n} \mathsf{Ed}(x, P_i)$, i.e., $c^* = \arg\min_{x \in L} \max_{i=1}^{n} \mathsf{Ed}(x, P_i)$, c^* is equal to the x-coordinate of the lowest point v^* in the upper envelope of $\mathcal{H}$. Therefore, to compute c^*, it is sufficient to find v^*.

As shown in [18] and mentioned in Section 1.2, in the discrete case where each uncertain point P_i has m discrete locations (each with a probability), each function $\mathsf{Ed}(x, P_i)$ consists of $m+1$ pieces

Download English Version:

https://daneshyari.com/en/article/1142143

Download Persian Version:

https://daneshyari.com/article/1142143

Daneshyari.com