ARTICLE IN PRESS

COMGEO:1525

Computational Geometry ••• (••••) •••–•••

Contents lists available at ScienceDirect

Computational Geometry: Theory and Applications

www.elsevier.com/locate/comgeo

On the separability of stochastic geometric objects, with applications ☆

Jie Xue, Yuan Li, Ravi Janardan

Department of Computer Science and Engineering, University of Minnesota-Twin Cities, Minneapolis, MN, 55455, USA

ARTICLE INFO

Article history:
Received 23 March 2017
Received in revised form 17 November 2017
Accepted 1 March 2018
Available online xxxx

Keywords:
Stochastic objects
Linear separability
Separable-probability
Expected separation-margin
Convex hull

ABSTRACT

In this paper, we study the linear separability problem for stochastic geometric objects under the well-known unipoint and multipoint uncertainty models. Let $S = S_R \cup S_B$ be a given set of stochastic bichromatic points, and define $n = \min\{|S_R|, |S_B|\}$ and $N = \max\{|S_R|, |S_B|\}$. We show that the *separable-probability* (SP) of S can be computed in $O(nN^{d-1})$ time for $d \geq 3$ and $O(\min\{nN \log N, N^2\})$ time for $d = 2$, while the *expected separation-margin* (ESM) of S can be computed in $O(nN^d)$ time for $d \geq 2$. In addition, we give an $\Omega(nN^{d-1})$ *witness-based lower bound* for computing SP, which implies the optimality of our algorithm among all those in this category. Also, a hardness result for computing ESM is given to show the difficulty of further improving our algorithm. As an extension, we generalize the same problems from points to general geometric objects, i.e., polytopes and/or balls, and extend our algorithms to solve the generalized SP and ESM problems in $O(nN^d)$ and $O(nN^{d+1})$ time, respectively. Finally, we present some applications of our algorithms to stochastic convex hull-related problems.

1. Introduction

Linear separability describes the property that a set of d-dimensional bichromatic (red and blue) points can be separated by a hyperplane such that all the red points lie on one side of the hyperplane and all the blue points lie on the other side. This problem has been well studied for years in computational geometry, and is widely used in machine learning and data mining for data classification. However, existing linear-separation algorithms require that all the input points must have fixed locations, which is rarely true in reality due to imprecise sampling from GPS, robotics sensors, or some other probabilistic systems. It is therefore essential to study the conventional linear separability problem under uncertainty.

In this paper, we study the linear separability problem under two different uncertainty models, i.e., the *unipoint* and *multipoint* models [1]. In the former, each stochastic data point has a fixed and known location, and has a positive probability to exist at that location; whereas in the latter, each stochastic data point occurs in one of discretely-many possible locations with a known probability, and the existence probabilities of each point sum up to at most 1 (to allow for its absence). Our focus is to compute the *separable-probability* (SP) and the *expected separation-margin* (ESM) for a given set of bichromatic stochastic points (or general geometric objects) in $\mathbb{R}^d$ for $d \geq 2$, where the former is the probability that the existent points (or objects) are linearly separable, and the latter is the expectation of the separation-margin of the existent points (or objects). (See Section 3.1 for a detailed and formal definition of the latter.)

☆ A preliminary version appeared at the *32nd International Symposium in Computational Geometry, 2016*.

E-mail addresses: xuexx193@umn.edu (J. Xue), lixx2100@umn.edu (Y. Li), janardan@umn.edu (R. Janardan).

https://doi.org/10.1016/j.comgeo.2018.06.001
0925-7721/© 2018 Published by Elsevier B.V.

Please cite this article in press as: J. Xue et al., On the separability of stochastic geometric objects, with applications, Comput. Geom. (2018), https://doi.org/10.1016/j.comgeo.2018.06.001

The brute-force approach of enumerating all possible instances takes exponential runtime. Therefore, in this paper, we propose novel algorithms that carefully compute SP and ESM in a much more efficient way. Furthermore, we show that our approach is highly extensible and can solve many other related problems defined on other types of objects or on multiple colors. To summarize, our main contributions are:

(1) We propose an $O(nN^{d-1})$-time algorithm, which uses linear space, for solving the SP problem when given a set of bichromatic stochastic points in $\mathbb{R}^d$, $d \geq 3$. (The runtime is $O(\min\{N^2, nN\log N\})$ for $d = 2$.) We also show an $\Omega(nN^{d-1})$ lower bound for all witness-based algorithms, which implies the optimality of our algorithm among all witness-based methods for $d \geq 3$. (See Section 2.)
(2) We show that the ESM of the above dataset can be computed in $O(nN^d)$ time for $d \geq 2$, using linear space. A hardness result is also given to show the total number of distinct possible separation-margins is $\Theta(nN^d)$, which implies that it may be difficult to achieve a better runtime. (See Section 3.)
(3) We extend our algorithms to compute the SP and the ESM for datasets containing general stochastic geometric objects, such as polytopes and/or balls. Our generalized algorithms solve the former problem in $O(nN^d)$ time, and the latter in $O(nN^{d+1})$ time, using linear space. (See Section 4.)
(4) We provide some applications of our algorithms to problems related to the stochastic convex hull (SCH). Specifically, by taking advantage of our SP algorithm, we give a method to compute the SCH membership probability, which matches the best known bound but is more direct. Also, we consider some generalized versions of this problem and show how to apply our separability algorithms to solve them efficiently. (See Section 5.)

1.1. Related work

The study of computational geometry problems under uncertainty is a relatively new topic, and has attracted a lot of attention; see [2] and [3] for two surveys. Different uncertainty models and related problems have been investigated in recent years. See [4–11] for example. The unipoint/multipoint uncertainty model, which we use in this paper, was first defined in [1,12], and has been applied in many recent papers. For instance, in [13], Kamousi et al. studied the stochastic minimum spanning tree problem, and computed its expected length. Suri et al. investigated the most likely convex hull problem over uncertain data in [12]; the convex hull was revisited by Agarwal et al. in [1], who showed how to compute the probability that a query point is inside the uncertain hull. In [14], Suri and Verbeek studied the most likely Voronoi Diagram (LVD) in $\mathbb{R}^1$ under the unipoint model, and the expected complexity of LVD was further improved by Li et al. in [15], who explored the stochastic line arrangement problem in $\mathbb{R}^2$. In [16], Agrawal et al. proposed efficient algorithms for the most likely skyline problem in $\mathbb{R}^2$ and gave NP-hardness results in higher dimensions. See [17–22] for more results under these uncertainty models.

Recently, in [7], de Berg et al. studied the separability problem given a set of bichromatic imprecise points in $\mathbb{R}^2$ in a setting that each point is drawn from an imprecision region.

Very recently, in [23], one of our proposed problems, the SP computing problem, was independently and simultaneously studied by Fink et al. under the same uncertainty model, and the same bounds were obtained, i.e., an $O(nN^{d-1})$)-time and $O(N)$-space algorithm was proposed to solve the problem in $\mathbb{R}^d$. However, in terms of techniques, Fink et al.'s method is very different from ours. Their computation of SP relies on an additional dummy anchor point and, based on this point, the probability is computed in an inclusion-exclusion manner. On the other hand, our method solves the problem more directly: it does not introduce any additional points and the SP is computed using a simple addition principle. Furthermore, our algorithm can be easily extended to solve many generalized problems (e.g., multiple colors, general geometric objects, etc.). Therefore, our solution for the SP computing problem is also of interest.

1.2. Basic notations and preliminaries

Throughout this paper, the basic notations we use are the following. We use $S = S_R \cup S_B$ to denote the given stochastic bichromatic dataset, where S_R (resp. S_B) is a set of red (resp. blue) stochastic points (or general geometric objects in Section 4). The notations n and N are used to denote the sizes of the smaller and larger classes of S respectively, i.e., $n = \min\{|S_R|, |S_B|\}$ and $N = \max\{|S_R|, |S_B|\}$, and d is used to denote the dimension. In this paper we always assume that d is a constant. When we need to denote a normal bichromatic dataset (without considering the existence probabilities), we usually use the notation $T = T_R \cup T_B$. The coordinates of a point $x \in \mathbb{R}^d$ are denoted as $x^{(1)}, x^{(2)}, \ldots, x^{(d)}$. If T is a dataset in $\mathbb{R}^d$ and U is some linear subspace of $\mathbb{R}^d$, we use T^U to denote a new dataset in the space U, which is obtained by orthogonally projecting T from $\mathbb{R}^d$ onto U.

We say a set of bichromatic points is *strongly separable* iff there exists a hyperplane, h, so that all the red points strictly lie on one side of h while all the blue points strictly lie on the other side. Also, we can define the concept of *weakly separable* similarly, except that we allow points to lie on the hyperplane. A hyperplane that strongly (resp., weakly) separates a set of bichromatic points is called a *strong* (resp., *weak*) *separator*. The following is a fundamental and well-known result which we will use in various places of this paper.

Please cite this article in press as: J. Xue et al., On the separability of stochastic geometric objects, with applications, Comput. Geom. (2018), https://doi.org/10.1016/j.comgeo.2018.06.001