Influence Based Voronoi Diagrams of Clusters

Ziyun Huang[1] Danny Z. Chen[2] Jinhui Xu[1]

[1] Department of Computer Science and Engineering
State University of New York at Buffalo
{ziyunhua, jinhui}@buffalo.edu
[2] Department of Computer Science and Engineering
University of Notre Dame
dchen@cse.nd.edu

1 Overview

In this paper, we study the following generalization of Voronoi diagram, called *Influence-based Voronoi Diagram (IVD)*: Given a set of (possibly overlapping) clusters $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ of points in space $\mathbb{R}^d$ for some fixed integer $d > 0$, construct a Voronoi diagram of $\mathcal{C}$ so that the Voronoi cell of each cluster $C_i \in \mathcal{C}$ is the union of points in the space which have a larger joint influence from C_i than any other cluster in $\mathcal{C}$, where the influence from a cluster C_i to a point $q \in \mathbb{R}^d$ is measured by a non-negative influence function $F(C_i, q)$. The cluster $C_i \in \mathcal{C}$ which has the largest joint influence on q is called the *Maximum Influence Cluster* (MIC) of q, and the process of finding MIC for a point (or a Voronoi cell) is called *Assignment*. There are two major differences between IVD and a traditional Voronoi diagram (VD). One difference is that each Voronoi site of IVD is a given point cluster, while the Voronoi site of a traditional VD is often one input point. The other difference is that IVD is influence based, while the traditional VDs are in general distance based, meaning that when determining the Voronoi cell for a point q, IVD measures the joint influence from all points in a Voronoi site to q by an influence function, but the traditional VDs mainly compute the distance (*e.g.,* Euclidean or Hausdorff distance) between the Voronoi site and q.

Chen *et al.* were the first ones to study joint influence in Voronoi diagram. In a recent work called *Clustering Induced Voronoi Diagrams (CIVD)* [3], they showed that for any set P of n points in $\mathbb{R}^d$ and an influence function $F(C, q)$ satisfying certain general conditions, it is possible to partition the $\mathbb{R}^d$ space into $O(n \log n)$ cells so that all points in each cell share a common subset of P as their $(1 - \epsilon)$-approximate MIC for some small constant $\epsilon > 0$. CIVD considers all possible subsets (*i.e.,* the power set) of P as its potential Voronoi sites, and the Voronoi sites, as well as the cells, are solely induced by the influence function. It is showed in [3] that a general technique called *Approximate Influence (AI)* decomposition exists for a large class of influence functions to partition the space, however it is quite unlikely to find a common assignment algorithm for such a class of influence functions (the assignment problem was solved in [3] in a problem-dependent manner).

In some sense, the IVD model can be viewed as a special case of CIVD by restricting the Voronoi sites to be the set of given clusters $\mathcal{C}$, instead of the power set of the input points. We expect that such a restriction can enable us to characterize a set of general conditions (which are slightly different from those of the CIVD model) such that for any influence function satisfying this set of conditions its corresponding IVD can be computed by a common algorithm, thus simultaneously solving the space partition and assignment problems.

The IVD model is motivated by a recent interesting application of joint influence in medicine. In [10, 11], Wang *et al.* showed that by constructing a Voronoi diagram of a collection of density-based clusters of already identified candidate neutrophils, it is possible to significantly improve the accuracy (by about 10%) of identifying true unidentified neutrophils in H&E staining histology tissue images. Their method suggests that an unidentified neutrophil in the Voronoi cell of a cluster of already identified neutrophils is more likely to be a true neutrophil. This provides a new way (*i.e.,* using the joint influence of a set of identified neutrophils to discover true neutrophils) to solve a challenging medical imaging problem.

It is worthy pointing out that several distance-based Voronoi diagrams allow their Voronoi sites to have multiple points. These include the k-th order Voronoi diagrams [8], the Hausdorff Voronoi diagrams

[9, 2, 13], and the two-point site Voronoi diagrams [1, 4–7, 12]. The distance functions used in these works are often defined by the closest (or farthest) point in the Voronoi sites, not by a collective effect of all points of these sites. The IVD problem studied in this paper is more general as the aforementioned problems can be viewed as its special cases.

To solve the IVD problem, our main idea is to utilize the CIVD technique's ability of handling influence functions for space partition. Particularly, we first treat the set $\mathcal{C}$ of clustered input points as a set P of unclustered points (*i.e.,* the union of all clusters) and use CIVD to partition the space for P. Then, we develop a divide-and-conquer algorithm to solve the assignment problem for a large class of influence functions. One challenge of such an approach is that the cells partitioned by CIVD are intended for clusters automatically induced by the influence function, which may be quite different from the given clusters in the input. Thus, the difficulty is how to resolve the inconsistency between the two types of clusters and still generates a small-size space partition for the IVD problem. To overcome this difficulty, we first extend the AI decomposition of CIVD to an Assisted AI decomposition. This ensures that the resulting space partition has certain property. This property is then used by the divide-and-conquer algorithm to recursively relate the cells of P to the cells of subsets of P. Such a relationship enables us to recursively resolve the assignment problem, and thus overcomes a main difficulty of CIVD. Consequently, we obtain a $(1-\epsilon)$-approximation algorithm for IVD which yields a Voronoi diagram of size $O(N \log N)$ in $O(T_2(N)N \log^2 N + T_1(N))$ time, where N is the total number of points in $\mathcal{C}$, $\epsilon > 0$ is a small constant, and $T_1(N)$ is the time needed for building a data structure for reporting a $(1 \pm \epsilon)$-approximate value of $F(C_i, q)$ in $T_2(N)$ time for any $C_i \in \mathcal{C}$ and $q \in \mathbb{R}^d$.

Since our technique does not need to know the exact form of the influence function (as long as it satisfies some general properties), we expect that it will be applicable to many applications.

References

1. G. Barequet, M.T. Dickerson, D. Eppstein, D. Hodorkovsky, and K. Vyatkina, "On 2-Site Voronoi Diagrams under Geometric Distance Functions," *Proc. 8th International Symp. on Voronoi Diagrams in Science and Engineering*, 2011, pp. 31-38.
2. P. Cheilaris, E. Khramtcova, S. Langerman, and E. Papadopoulou, "A Randomized Incremental Approach for the Hausdorff Voronoi Diagram of Non-crossing Clusters." *LATIN 2014*, pp. 96-107, 2014.
3. Danny Z. Chen, Ziyun Huang, Yangwei Liu and Jinhui Xu, "On Clustering Induced Voronoi Diagrams," *Proc. 54th Annu. IEEE Sympos. Found. Comput. Sci. (FOCS 2013)*, pp. 390-399, 2013.
4. M.T. Dickerson and D. Eppstein, "Animating a Continuous Family of Two-site Voronoi Diagrams (and a Proof of a Bound on the Number of Regions)," *Proc. 25th ACM Symp. Computational Geometry*, 2009, pp. 92-93.
5. M.T. Dickerson and M.T. Goodrich, "Two-site Voronoi Diagrams in Geographic Networks", *Proc. 16th ACM SIGSPATIAL International Conf. Advances in Geographic Information Systems*, 2008.
6. I. Hanniel and G. Barequet, "On the Triangle-Perimeter Two-site Voronoi Diagram," *Trans. on Computational Science*, 9(2010), 54-75. doi:10.1145/1463434.1463504.
7. D. Hodorkovsky, "2-Point Site Voronoi Diagrams," M.Sc. Thesis, Technion, Haifa, Israel, 2005.
8. A. Okabe, B. Boots, K. Sugihara, and S.N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd Eds., John Wiley & Sons, 2000.
9. E. Papadopoulou, "The Hausdorff Voronoi Diagram of Point Clusters in the Plane," *Algorithmica*, 40(2004), 63-82.
10. J. Wang, J.D. MacKenzie, R. Ramachandran, and D.Z. Chen, "Identifying Neutrophils in H&E Staining Histology Tissue Images," *Proc. of the 17th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Part I,* 2014, pp. 73–80.
11. J. Wang, J.D. MacKenzie, R. Ramachandran, and D.Z. Chen, "Neutrophils Identification by Deep Learning and Voronoi Diagram of Clusters," accepted to the *18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI),* October 2015.
12. K. Vyatkina and G. Barequet, "On 2-Site Voronoi Diagrams under Arithmetic Combinations of Point-to-Point Distances," *Proc. 7th International Symp. Voronoi Diagrams in Science and Engineering*, 2010, pp. 33-41.
13. J. Xu, L. Xu, and E. Papadopoulou, "Computing the Map of Geometric Minimal Cuts." *Algorithmica*, 68(4): 805-834 (2014).