1. Name: Guang Xu
2. Email: guangxu@cse.buffalo.edu
3. Dissertation Title: Approximation Algorithms for Clustering and Related Problems
4. Committee Chair: Dr. Jinhui Xu
5. Committee Members: Dr. Xin He and Dr. Hung Quang Ngo

Abstract

The clustering problem is a well-studied problem in computer science. As an unsupervised learning procedure, a number of variants of clustering problem have been investigated and many important approaches have been proposed. Among these variants, the Euclidean k-center problem and Euclidean k-median problem have received considerable attentions from the fields of computational geometry and combinatorial optimizations. Such problems are NP-hard, and have been studied from the perspective of worst case performance guarantees by exploring various approximation algorithm design techniques in this dissertation.

First, we study the clustering problem with prescribed grouping information. In such clustering problem, some pre-existing partial grouping information, this requires that some points must belong to the same cluster, has to be preserved by the clustering algorithm. Our main results about clustering problem with prescribed grouping information can be summarized as follows: 1.An O(m + n log k)-time 3-approximation algorithm for the KCWPGI problem. 2.An O(km)-time 2.73-approximation algorithm for the KCWPGI problem. Next, we study clustering problem on bipartite graph formed by a set of facility locations and client locations, which is also called facility location problem. Our results for the UFLWP problem can be summarized as follows: 1. An LP rounding based 2.7-approximation algorithm for UFLWP problem. 2. A combinatorial 1.85-approximation algorithm for UFLWP problem.

Then, we apply the idea of grouping "similar" things to tackle some NP-hard problems which are seemingly far away from clustering problem. One of such problem is rectangle stabbing problems. Several efficient constant approximation algorithms have been developed in this dissertation. Finally, we design some practically efficient algorithms for biplane imaging geometry determining problem.