Geometry Seminar
November 21, 2006, Tuesday, 6:00 p.m.
Room 613, Courant Institute
251 Mercer Street, New York

Linear-space implementation of multi-dimensional scaling with applications

Tetsuo Asano
JAIST, Kanazawa, Japan

Abstract

Suppose a set of n objects is given and for each pair of objects (i, j) their dissimilarity denoted by $\delta_{i,j}$ can be computed in an efficient way, say constant time. Using the dissimilarity information, we want to map objects into points in some low dimensional space while dissimilarities are preserved as distances between corresponding points. We encounter this kind of situations often in practice. Converting distance information into coordinate information is helpful for human perception because we can see how close two objects are. Because of practical importance, this topic has been widely studied also under the name of dimension reduction. Multi-Dimensional Scaling (MDS) is a generic name for a family of algorithms doing such conversion from distances to coordinates. Although MDS is powerful, it has a serious drawback for practical use, that is, its high space complexity. Input to MDS is a n times n matrix specifying pairwise dissimilarities (or distances). Fortunately, we can avoid this high space complexity if dissimilarity information is given by a function to evaluate dissimilarity in an efficient way, say each in constant time. If it must be specified by a matrix, we cannot avoid quadratic space. In this paper we propose a scheme for implementing MDS using only linear working space without losing power of the original MDS.

A key idea for linear-space implementation is to use clustering in the first step. The quality of the output heavily depends on clustering. In many applications, dissimilarities between similar objects have more importance than those between totally dissimilar objects. We propose a simple algorithm for finding a size- constrained clustering that achieves largest inter-cluster distance, or maximizes the smallest distance between objects from different clusters. We also present some interesting results on hardness of problems discussed above. We formulate the problem as a distance preserving graph embedding problem. We show there are classes of graphs for which the corresponding problem is NP-hard while the problem associated with a complete graph can be solved in polynomial time using MDS. Joint work with Prosenjit Bose, Paz Carmi, Mathieu Couture, Anil Maheshwari, Chang Shu, Michiel Smid, and Stefanie Wuhrer.

For further information contact {pach,pollack}@cims.nyu.edu, or visit our website: http://www.math.nyu.edu/seminars/geometry_seminar.html