

MULTIPLE PRINCIPAL COMPONENT ANALYSIS ON HIGH DIMENSIONAL DATA

Kittisak Kerdprasop, and Nittaya Kerdprasop

Data Engineering and Knowledge Discovery (DEKD) Research Unit,
School of Computer Engineering, Institute of Engineering,
Suranaree University of Technology

Abstract

Clustering is a widely used technique to discover homogeneous groups, or clusters, of data according to a certain similarity measure. Many algorithms have been designed to compute a partition on full-dimensional data set. While these approaches work successfully on low-dimensional data sets, their efficiency decrease significantly in higher dimensional space. In high dimensional data, some dimensions tend to be redundant or irrelevant. Massive dimensions can confuse the clustering algorithms. It is also difficult to group similar data points in very high dimensions because the distance between any two data points becomes almost the same. The most difficult problem on clustering high dimensional data is that different clusters may exist in different subspaces of different dimensions. Therefore, we propose a technique of multiple principle component analysis to cope with the difficulty of clustering high dimensional data.

ตีพิมพ์ใน: Proceedings of the Research Network Development of Higher Education Alliance in Nakhon Ratchasima, Suranaree University of Technology, Nakhon Ratchasima, Thailand, June 24, 2005. pp.50-51.