Abstract

Data matrix having the same set of entity in the rows and columns is known as one-mode data matrix, and traditional one-mode clustering algorithms can be used to cluster the rows (or columns) separately. With the popularity of use of two-mode data matrices where the rows and columns have different sets of entities, the need for simultaneous clustering of rows and columns popularly known as two-mode clustering increased. Additionally, the emergence of large data sets and the prediction of Moore’s law slow-down have created the challenge of clustering scalability. In this paper, we address the problem of scalability of organizing an unlabelled two-mode dataset into clusters utilizing multicore processor. We propose a parallel genetic algorithm (GA) heuristics based two-mode clustering algorithm, which is an adaptation of the classical Cuthill-McKee Matrix Bandwidth Minimization (MBM) algorithm. The classical MBM method aims at reducing the bandwidth of a sparse symmetric matrix, which we adapted to make it suitable for non-symmetric real-valued matrix. Preliminary results indicate that our algorithm is scalable on multicore processor compared to serial implementation. Future work will include more extensive experiments and evaluations of the system.
Scalability of Parallel Genetic Algorithm for Two-mode Clustering


Scalability of Parallel Genetic Algorithm for Two-mode Clustering

edu/biclustering/yeast. matrix

**Index Terms**

Computer Science
Artificial Intelligence

**Keywords**

Scalability; two-mode clustering; parallel genetic algorithm; matrix reordering