

Gene Ordering in Partitive Clustering using Microarray Expressions

Shubhra Sankar Ray¹, Sanghamitra Bandyopadhyay², and Sankar K. Pal¹

¹Center for Soft Computing Research: A National Facility, ²Machine Intelligence Unit,
Indian Statistical Institute,

Email: {shubhra_r, sanghami, sankar}@isical.ac.in

Abstract

Motivation: A central step in the analysis of gene expression data is the identification of groups of genes that exhibit similar expression patterns. Clustering and ordering the genes using gene expression data into homogeneous groups was shown to be useful in functional annotation, tissue classification, regulatory motif identification, and other applications. Although there is a rich literature on gene ordering in hierarchical clustering framework for gene expression analysis, to the best knowledge of the author, there is no work addressing and evaluating the importance of gene ordering in partitive clustering framework. Outside the framework of hierarchical clustering, different gene ordering algorithms are applied on the whole data set, and the domain of partitive clustering is still unexplored with gene ordering approaches.

Results: A new hybrid method for ordering genes in each of the clusters of partitive clustering solution using microarray gene expressions is proposed. Two existing ordering algorithms used for optimally ordering cities in Travelling Salesman Problem (namely, Concorde [1] and FRAG_GA [2]), are hybridized individually with Self Organizing MAP to show the importance of gene ordering in partitive clustering framework. We validated our hybrid approach using Yeast and Fibroblast data and showed that our approach improves the result quality of partitive clustering solution, by identifying subclusters within big clusters, minimization of summation of gene expression distances and the maximization of biological gene ordering using MIPS categorization. Moreover, the new hybrid approach, finds comparable or sometimes superior biological gene order than those obtained by optimal leaf ordering in hierarchical clustering solution [3].

Reference:

[1] D. Applegate, R. Bixby, V. Chvátal, and William Cook, "Concorde Package. [Online]," www.tsp.gatech.edu/concorde/downloads/codes/src/co031219.tgz, 2003.

[2] Shubhra Sankar Ray, Sanghamitra Bandyopadhyay and Sankar K. Pal, "Genetic Operators for Combinatorial Optimization in TSP and Microarray Gene Ordering", *Applied Intelligence* (published online: epub ahead of print), 2006

[3] Z. Bar-Joseph, D. K. Gifford, and T. S. Jaakkola, "Fast optimal leaf ordering for hierarchical clustering," *Bioinformatics*, vol.-17, pp- 22-29, 2001.