



TECHNOLOGY

Hierarchical Clustering Explorer (HCE)

OVERVIEW

Multidimensional data sets are common in many research areas, including microarray experiment data sets. Genome researchers are using cluster analysis to find meaningful groups in microarray data. Some clustering algorithms, such as k-means, require users to specify the number of clusters as an input, but users rarely know the right number beforehand. Other clustering algorithms automatically determine the right number of clusters, but users may not be convinced of the result since they had little or no control over the clustering process. To avoid this dilemma, the Hierarchical Clustering Explorer (HCE) applies the hierarchical clustering algorithm without a predetermined number of clusters, and then enables users to determine the natural grouping with interactive visual feedback (dendrogram and color mosaic) and dynamic query controls. HCE 1.0 implemented four general techniques that could be used in interactive explorations of clustering results.

- * overview of the entire dataset, coupled with a detail view so that high-level patterns and hot spots can be easily found and examined
- * dynamic query controls so that users can restrict the number of clusters they view at a time and show those clusters more clearly
- * coordinated displays: the overview mosaic has a bi-directional link to 2-dimensional scatterplots
- * cluster comparisons to allow researchers to see how different clustering algorithms group the genes.

However, the high dimensionality of the data sets still hinders users from finding interesting patterns, clusters, and outliers. Determining the biological significance of such features remains problematic due to the difficulties of integrating biological knowledge. In addition, it is not efficient to perform a cluster analysis over the whole data set in cases where researchers know the approximate temporal pattern of the gene expression that they are seeking. To address these problems, we developed the Hierarchical Clustering Explorer 2.0 by adding three new features to HCE:

- * scatterplot ordering methods so that all 2D projections of a high dimensional data set can be ordered according to relevant criteria
- * a gene ontology browser, coupled with clustering results so that known gene functions within a cluster can be easily studied
- * a profile search so that genes with a certain temporal pattern can be easily identified.

For further information, please contact the Office of Technology Commercialization, (301) 405-3947 E-mail: otc@umd.edu.

CONTACT INFO

UM Ventures
0134 Lee Building
7809 Regents Drive
College Park, MD 20742
Email: umdtechtransfer@umd.edu
Phone: (301) 405-3947 | Fax: (301) 314-9502

Additional Information

INSTITUTION

University of Maryland, College Park

LICENSE STATUS

Available for non-exclusive license

CATEGORIES

- Information Technology

EXTERNAL RESOURCES

IS-2003-068