

Visual and Analytical Tools for Cluster Analysis

Urska Cvek

Department of Computer Science and Laboratory for Advanced Biomedical Informatics, LSU-Shreveport, LA; Department of Bioinformatics and Computational Biology, Bioinformatics Core, LSUHSC-Shreveport, LA

Background and Objective

The output of a clustering algorithm is a list of clusters and their membership. With the number of clustering algorithms available the problem becomes the ability to analyze multiple clustering methods and their results. The goal of the project is to design new analytical and visual tools and techniques that provide insight into these multiple clustering algorithm results.

Methods

The basic tools enable visualizations (scatterplots, parallel coordinates, radviz, etc.) and data analyses (clustering, PCA, etc.). Our measure detects clusters, creates a heatmap, hierarchical tree, treemap or graph. The algorithm clusters data into meta-groups and determines the largest number of clusters that a record shares with others.

Results

We formally define similarity and dissimilarity of clustering results. We extend the two-dimensional displays into the third dimension using a number of indices and measures for inter- and intra- cluster tightness through a set of interaction methods.

Discussion and Conclusions

We show record, cluster size comparison, proximity analysis, cluster consolidation and representation in 2D & 3D on a biomedical data set.

Theme: Biotechnology and Bioinformatics

The project was supported by Grant Number P20RR016456 from the National Center For Research Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center For Research Resources or the National Institutes of Health.