Introduction

Very often scientific data are defined by multidimensional vectors of numerical values. To enable exploratory data analysis involving heuristic abilities of human experts visualization of data is highly desirable: a picture is worth a thousand words.

There are different approaches to visualization. We consider one of the most popular approaches known as multidimensional scaling (see for example Borg I, Groenen P (2005) *Modern Multidimensional Scaling*. 2nd ed. Springer, New York). By means of this technique a set of multidimensional vectors can be represented as a set of points in a low-dimensional space, and exposed in this way to a human expert for heuristic analysis. Application areas of multidimensional scaling vary from psychometrics and market analysis to mobile communication, data mining, pharmacology and biomedicine.

An essential part of the technique is optimization of a function possessing many optimization adverse properties. An image of a set of objects is sought as a set of points in a low-dimensional metric embedding space with pairwise distances between the image points fitting the corresponding given dissimilarities.

The problem of construction of images of the considered objects is reduced to minimization of an accuracy of fit criterion, e.g. of a least squares Stress function. Stress is not everywhere differentiable. The function normally has many local minima. It is invariant with respect to translation, rotation and mirroring. The minimization problem of Stress is high dimensional. Therefore minimization of Stress function is a difficult global optimization problem. Various global optimization algorithms are used in this project.

Applications

Within the framework of BalticGrid projects the group of researchers at the Institute of Mathematics and Informatics has gridified several two level global optimization algorithms for multidimensional scaling based on stochastic and deterministic global and local optimization.

Input files contain real numbers defining dissimilarities.
Output files contain coordinate numbers of image points.
Post-processing is used for visualization.
An example of application: Identification of the heart rate parameters relevant to characterization of sleep stages. Several linear (mean, standard deviation, spectral parameters) and nonlinear (approximate entropy, scaling exponent of detrended fluctuation analysis, slope of a curve of progressive detrended fluctuation analysis) parameters of heart rate have been estimated for each time period of sleep. Relevance of the considered parameters with respect to characterization of sleep stages is analyzed using multidimensional scaling.

The main limiting resource for the applications of multidimensional scaling is computing time. Depending on the data the runtime of algorithms is from seconds to days and weeks. Some problems are not solved with acceptable accuracy in acceptable time. Therefore applications benefit from the Grid environment which provides powerful computing resources.

Main publications of the research group on this research topic:


Nearest goals
Future plans are to develop new global optimization algorithms for multidimensional scaling, implement efficient parallel versions, investigate their efficiency and apply them to solve practical problems.

Contact info:

http://sig.balticgrid.org/SIGs/multidimensional-scaling/
Julius Žilinskas zilinskasjulius@gmail.com
Institute of Mathematics and Informatics http://www.mii.lt