

Title:

Introduction to bioinformatics data sets mining using fuzzy biclustering

Organizers:

Francesco Masulli (University of Genova & Temple University)

Stefano Rovetta (University of Genova)

Abstract: The analysis of genomic data from DNA microarray can produce a valuable information on the biological relevance of genes and correlations among them. In the last few years some biclustering techniques have been proposed and applied to this analysis. Biclustering is an un-supervised learning task aimed to find clusters of samples possessing similar characteristics together with features creating these similarities. Starting from the seminal paper by Cheng and Church published in 2000, many biclustering techniques have been proposed for bioinformatic data analysis. It is especially useful when applied to the analysis of DNA microarray data since it can tackle the important problem of identifying genes with similar behavior with respect to different conditions. Some biological tasks where biclustering can be successfully applied are: (1) Identification of co-regulated genes and/or specific regulation processes; (2) Gene functional annotation; (3) Sample and/ or tissue classification. In this tutorial we will focus on the fuzzy model of biclustering as it is very promising from both a computational and a representation point of view. This model allows finding multiple solutions (thus avoiding problems such as random) with significant speed. Moreover, some techniques, based on the fuzzy possibilistic approach to clustering, can find very large and homogeneous biclusters, as shown by experimental results. In the tutorial we will present also an experimental assessment of fuzzy biclustering algorithms, using some computationally parsimonious stability indexes.