



UMass Lowell Computer Science Colloquium Announcement

Speaker: Dr. T. M. Murali
Center for Advanced Genomic Technology
Bioinformatics Program, Boston University

Date & Time: March 10, 2003, 3:00pm-4:00pm

Place: Olsen 311. **Refreshments are served at 2:45pm**

Projective Clustering of High-Dimensional Data

We present projective clustering, a new technique for clustering points in high dimensional spaces. A projective cluster of a set of points is a subset of the points and a subset of the dimensions with the property that along each selected dimension, the coordinates of all the selected points lie within a small interval; thus, the selected points cluster well only when projected into the selected dimensions.

Projective clustering has three principal advantages. First, we can apply it to data sets that contain clusters only in subspaces of the original space, a situation that "full-dimensional" clustering algorithms like k-means cannot handle. Second, we can apply it to data sets where different subsets of points cluster in different lower-dimensional subspaces. In contrast, a clustering technique that first employs a dimensionality-reduction technique such as PCA would choose a single lower-dimensional subspace for all the clusters. Finally, since the dimensions in each projective cluster are a subset of the original dimensions, our technique is ideal in situations when the dimensions have meaning, e.g., in biological data. Thus, projective clustering combines the advantages of popular clustering and dimensionality reduction techniques, while avoiding some of the drawbacks of these methods.

We describe a probabilistic algorithm for computing "large" projective clusters and present bounds on the running time of our algorithm. We have applied our algorithm to find xMotifs or conserved gene expression motifs in DNA microarray data. We present results that demonstrate that our technique finds biologically-meaningful patterns of co-expression.