APPLIED MATHEMATICS COLLOQUIUM

FUNCTIONAL CLUSTERING OF TEMPORAL MICROARRAY DATA

PING MA
Harvard University

ABSTRACT:

Microarray technology enables the scientist to measure the mRNA expression levels of thousands of genes simultaneously. When a series of micorarray experiments are conducted sequentially during a biological process, we call the resulting dataset a "temporal microarray dataset," which can provide insights on the underlying biology and help decipher the dynamic gene regulatory network. Clustering genes with similar temporal profiles is a crucial first step to reveal potential relationships among the genes.

In this talk, I present a functional clustering method based on a mixture smoothing-spline model. For each cluster, we model its mean profile using a smoothing spline and describe its individual gene's variation by a parametric random effect. We present an EM algorithm to find the maximum a posteriori. Our method automatically takes care of the missing data and infers the number of clusters in the data. Using the method, we analyzed a microarray dataset consisting of the results from 69 individual microarray experiments conducted over the life cycle of D. melanogaster. The resulting clusters were validated by examining the statistical over-representation of certain biological functions using the Gene Ontology database. The majority of clusters we had obtained are enriched for known and expected biological functions.

This is joint work with Cristian Castillo-Davis, Wenxuan Zhong, and Jun Liu.

MONDAY, FEBRUARY 14, 2005
4:15 PM
Building 4, Room 231

Refreshments at 3:30 PM in Building 2, Room 349.

Applied Math Colloquium: http://www-math.mit.edu/amc/spring05
Math Department: http://www-math.mit.edu