Electrical Engineering Colloquium
Dallas Chapter of IEEE Signal Processing Society and Dallas Chapter of IEEE Engineering in Medicine and Biology Society Present

Biclustering and the Search for Group Biomarkers

Professor Ahmed Tewfik, University of Minnesota

11am, Wednesday, December 5, 2007, ECSN 2.112

One of the major goals of gene expression data analysis is to uncover genetic pathways, i.e., chains of genetic interactions. For example, a researcher may be interested in identifying the genes that contribute to a disease. This task is difficult because subgroups of genes display similar activation patterns *only* under certain experimental conditions. Genes that are co-regulated or co-expressed under a subset of conditions will behave differently under other conditions. Finding genetic pathways therefore could be aided by identifying clusters of genes that are co-expressed under subsets of conditions as opposed to all conditions. A high degree of correlation between the activity levels of subsets of genes under subsets of conditions does not of course necessarily imply causality relations. Further biological analysis would be required to find actual genetic pathways. In this talk, I will describe how data is generated to find genes that are co-regulated. I discuss problems associated with microarray data collection, such as noise, noise characterization and missing data. I focus in particular on the selection of material, using the detection of ovarian cancer markers as an example. Next, I will provide an overview of biclustering techniques and their applications in genomics. Biclustering is also known as bidimensional clustering, subspace clustering and co-clustering in other application fields. Biclustering techniques produce local models whereas clustering approaches compute global models. If we use a clustering algorithm on the rows of the gene expression matrix, a given gene cluster is defined using all the conditions. In contrast, a biclustering technique will assign a gene to a bicluster based on a subset of conditions. Furthermore, when a clustering algorithm is applied to the rows of the gene expression matrix, it assigns each gene to a single cluster. Biclustering techniques on the other hand identify clusters that are not mutually exclusive or exhaustive. A gene may belong to no cluster, one or more clusters. By simultaneously clustering the rows and columns of the gene expression matrix, one can identify candidate subsets of conditions that may be associated with cellular processes that exhibit themselves only or subsets of genes that potentially play a role in a given biological process.

Ahmed H. Tewfik is the E.F. Johnson Professor of Electronic Communications with the Department of Electrical Engineering at the University of Minnesota. His current research interests are in genomics and proteomics, wearable sensors for patients at cardiac risk or with traumatic brain injury, programmable wireless networks, brain computing interfaces, healthcare safety and datanomic and pervasive computing and storage. Prof. Tewfik is a Fellow of the IEEE. He was a Distinguished Lecturer of the IEEE Signal Processing Society in 1997 - 1999. He received the IEEE third Millennium award in 2000. He was elected to the board of governors of the IEEE Signal Processing Society in 2005. He was selected to be the first Editor-in-Chief of the IEEE Signal Processing Letters from 1993 to 1999. He is currently an Associate Editor of the EURASIP Journal on Bioinformatics and Systems Biology.

For more information on the Dallas Chapter and directions to UTD, please refer to
http://www.utdallas.edu/~kehtar/ieee-sp