

Data Mining Techniques for Microarray Datasets

Lei Liu

U. of Illinois, Urbana-Champaign
leiliu@uiuc.edu

Jiong Yang

Case Western Reserve U.
jjong@eecs.cwru.edu

Anthony K. H. Tung

National U. of Singapore
atung@comp.nus.edu.sg

Development in microarray technology has result in revolutionary changes in biological research. Using microarrays, the expression level for thousands of genes can be monitored simultaneously, providing biologists with new ways to gain insight into the complex interaction in living organisms. To do so however, biologists must first overcome the challenge involved in analyzing the large and complex datasets that are generated from microarray experiments. Data mining research, which focuses on scalable and effective knowledge discovery from databases, can provide timely solutions for the biologists in these aspects. In this proposed seminar, we aim to provide platform in which various aspects of microarray data analysis will be introduced. In the first part of the seminar, we will discuss in layman term how microarray datasets are generated and used in biological research. We will use example from the real projects that we participate in to illustrate the potential of different technologies. In the second part of the tutorial, we will discuss existing data mining tools and methods used for analyzing the microarray data sets and their biological implications. Finally, we will present a set of open problems and future research directions for microarray data analysis.

Our goal is to make this tutorial a practical guide for microarray data analysis while at the same time to highlight some of the interesting research issues that arise in mining microarray datasets. Participants without any exposure to bioinformatics will be able to learn the basic concepts and techniques of microarray analysis. For those who are familiar with the technologies, we will offer a wide range of analysis tools that can be applied to microarray gene expression analysis. For people that would like to start research in this area, we would present a set of open problems and potential research directions. Previous exposure to some basic knowledge of molecular biology will be helpful, but not required for the tutorial. But we will assume audience's familiarity with basic data mining concepts.

Biography

As the founding director of the bioinformatics unit, **Dr. Lei Liu** joined the W. M. Keck Center for Comparative and Functional Genomics in 1999. Prior to coming to the University of Illinois, he worked as a postdoctoral fellow

for two years at the Department of Computer Science and Engineering at the University of Connecticut, where he also received a Ph.D. in cell biology. His expertise is in the areas of comparative genomics, biological databases, and data mining. He has been working in the microarray analysis and data mining for more than three years and co-authored several papers in that area. He has been organizing and participating in many workshops on microarray analysis at the University of Illinois. He has participated recently in the international workshop on "Statistical Methods in Microarray Analysis" in Singapore and presented a talk on multiple platform comparison. He collaborates with computer scientists and statisticians on developing new algorithm for microarray data mining.

Dr. Jiong Yang currently is a Schroeder Assistant Professor of the EECS department at Case Western Reserve University. He received his master and Ph.D. degree in Computer Science from UCLA at 1996 and 1999, respectively. He has been working on mining biological data in the past several years. Recently, he has authored and co-authored several publications in various database, data mining, and bioinformatics conferences and journals on the topic of mining microarray data sets. He has worked on frequent pattern discovery, clustering, and classification algorithms on various microarray data sets. He is an instructor of the course "data mining on bioinformatics" which was offered at UIUC. He recently has organized a workshop on Data Mining in Bioinformatics and he is a guest editor fro the TKDE special issue on Mining Biological Data.

Dr. Anthony K. H. Tung is currently an Assistant Professor in the Department of Computer Science, National University of Singapore (NUS). He received both his B.Sc. and M.Sc. in computer sciences from the National University of Singapore in 1997 and 1998 respectively. In 2001, he received the Ph.D. in computer sciences from Simon Fraser University (SFU). His research interests involve various aspects of databases and data mining (KDD) including buffer management, frequent pattern discovery, spatial clustering, outlier detection and classification analysis. Recent interest also includes data mining for microarray data and 3D protein structures, spatial indexing and sequences searches.