

Topic : **Bioinformatics and Machine Learning**

Speaker : Sun-Yuan Kung (貢三元), IEEE Fellow
Princeton U., USA

Date : 06/06/2008, 14:00 ~ 15:30

Place : F6011, Dept. of EE, National Sun Yat-Sen University

Abstract

Genomic bioinformatics represents a natural convergence of life sciences, computer sciences, and device technologies. The genome is not just a collection of genes working in isolation, rather it encompasses the global and highly coordinated control of information to carry out a range of cellular functions. The genome-wide study is vital for identification of correlated genes and understanding the mechanisms underlying gene transcription and regulation. One of the key challenges lies in the revelation of rich genomic information for the purpose of classification/prediction of cancers and tracking responses to drug therapies.

Genomic sequencing and gene expression technologies represent two vital fronts to drug design and disease classification. Even though recent progress has shown that, profile-based methods (i.e. profile-to-profile comparisons) can yield more sensitive and accurate comparisons of distantly related protein sequences, the sequencing approach still proves computationally prohibitive for a genome-wide study. In contrast, a collection of microarray experiments can yield a data matrix, whose rows and columns respectively reveal the expression level of tens of thousands of genes, and a number of independent or time-course conditions. More recently, fusion of both gene sequences and expression profiles appears to offer a promising new approach to the genomic study.

The success of bioinformatics hinges upon the imperative partnership between the genomic science and machine learning. Traditional machine learning tools need to be revisited and their algorithms must be specially tailored to fit the genomic data. For example, there are two machine learning subsystems: (1) feature extraction and (2) adaptive cluster discovery, the former involves how to correctly look at the genomic data while the latter how to classify them. It has now become evident that the feature extraction is by far the most critical to the success of genomic data mining.

This talk will discuss the fundamental roles machine learning may have on the advances in bioinformatics, including :

(i) Understanding of bioinformatic data and underlying models;

(ii) Feature extraction from bioinformatic data;

(iii) Machine learning techniques for bioinformatic data mining;

(iv) Case studies on real genomic data (both public and proprietary) on protein sequences and microarray data.

Sponsored by :

- Dept. of Electrical Engineering, National Sun Yat-Sen University
- IEEE Circuits & Systems Society (CAS) Tainan Chapter
- IEEE Solid-State Circuits Society (CAS) Tainan Chapter
- Engineering Technology Research & Promotion Center (ETRPC), National Sun Yat-Sen University