

THE UNIVERSITY



OF HONG KONG

Department of Mathematics

Numerical Mathematics and Applied Analysis Group
NMAA Half-day Workshop on Bioinformatics

Aug 25, 2011 (Thursday)
Room 309, Run Run Shaw Bldg, HKU

-
- | | |
|---------------|---|
| 9:50 – 10:10 | Miss Xi Chen, The University of Hong Kong
<i>Finding Optimal Control Policy by Using Dynamic Programming in Conjunction with State Reduction</i> |
| 10:10 – 10:30 | Miss Hao Jiang, The University of Hong Kong
<i>Kernel Methods for Biological Applications</i> |
| 10:30 – 11:30 | Professor Kevin Yuk-Lap Yip, The Chinese University of Hong Kong
<i>Integrating Experimental and Computational Data in Genome-wide Analysis</i> |
| 11:30 – 12:30 | Professor Shanfeng Zhu, Fudan University
<i>Efficient Semi-Supervised MEDLINE Document Clustering</i> |

All are welcome



Department of Mathematics

Numerical Mathematics and Applied Analysis Group**NMAA Half-day Workshop on Bioinformatics****Aug 25, 2011 (Thursday)****Room 309, Run Run Shaw Bldg, HKU****Miss Xi Chen**

The University of Hong Kong

**Finding Optimal Control Policy by
Using Dynamic Programming in Conjunction with State Reduction****9:50 - 10:10****Abstract**

We study the problem of finding optimal control policy for probabilistic Boolean networks (PBNs). Previous works have been done by using dynamic programming-based (DP) method. However, due to the high computational complexity of PBNs, DP method is computationally inefficient for large networks. Inspired by the state reduction strategies studied in [Qian04], we consider using dynamic programming in conjunction with state reduction approach to reduce the computational cost of DP method. Numerical examples are given to demonstrate the efficiency of our proposed method.

Miss Hao Jiang

The University of Hong Kong

Kernel Methods for Biological Applications**10:10 - 10:30****Abstract**

Kernel methods (KMs) are a class of algorithms for pattern analysis, whose best known element is the support vector machine (SVM). We consider the classification problem of biological data sets which is an important and hot topic in bioinformatics. In protein classification, we propose a novel kernel based on the K-Spectrum Kernel by incorporating physico-chemical and biological properties of amino acids as well as the motif information for the captured protein classification problem. We adopt the Eigen-matrix translation techniques for improving the classification accuracy. Experimental results indicate that the string-based kernel in conjunction with SVM classifier performs significantly better than the traditional spectrum kernel method. For ageing research, intensive research has been carried out to elucidate the role of DNA repair systems in the ageing process. We develop a linearly combined kernel with Support Vector Machine (SVM) to analyze the ageing related data. The popular supervised learning algorithm enables better discrimination between ageing-related and non-ageing-related DNA repair genes. Through training on the whole data set, we can identify the same important genes that target essential pathways as well. What's more, novel genes are detected which may reveal possible insights for biologists in ageing research.

All are welcome



Department of Mathematics

Numerical Mathematics and Applied Analysis Group**NMAA Half-day Workshop on Bioinformatics****Aug 25, 2011 (Thursday)****Room 309, Run Run Shaw Bldg, HKU****Professor Kevin Yuk-Lap Yip**

The Chinese University of Hong Kong

Integrating Experimental and Computational Data in Genome-wide Analysis**10:30 - 11:30****Abstract**

Recent advances in sequencing technology have enabled systematic surveying of molecular activities at the genome scale. Retrieving useful information from the vast amount of data produced is currently a big challenge. In this talk we describe our efforts in two recent studies in mining large-scale genomic data by integrating different types of experimental and computational features. In the first study, we set out to systematically identify all non-coding RNAs (ncRNAs) in *C. elegans*, which are believed to play important roles in gene regulation but have not been fully cataloged. In the second study, we investigated the relationships between chromatin features such as histone modifications and gene expression. Both studies reveal intricate interactions between different features, and demonstrate the power of data integration in genomic analysis.

Professor Shanfeng Zhu

Fudan University

Efficient Semi-Supervised MEDLINE Document Clustering**11:30 - 12:30****Abstract**

Combining multiple information for biomedical document clustering is a subject of intense research. For example, recently the performance of document clustering was enhanced by using both content and MeSH (Medical Subject Heading) semantic information, which were however linearly combined. The simple linear combination could be ineffective, because its representation space is too limited to combine multiple information sources considering the difference in their reliability. To relax this problem, we propose a new semi-supervised clustering method, SSNCut, which incorporates positive (must-link) and negative (cannot-link) constraints in terms of the cost function of spectral learning with normalized cut. We apply SSNCut to MEDLINE document clustering, reasonably assuming that document pairs with high semantic similarities have positive constraints to be in the same cluster and those with low similarities have negative constraints to be in different clusters. Experimental results with various 100 datasets of MEDLINE records show that SSNCut outperformed a linear combination method, being statistically significant.

All are welcome
