
Computer Science Colloquium

Machine Learning Research in Prof. Livingston's Group

Patrick Shaughnessy, Thao Nguyen, Chun-Yin (Cathy) Liu
Graduate Students, Computer Science Dept, UMass Lowell

Wednesday, 2 March 2005

Olsen 311

Refreshments at 2:30, Talk from 3:00-4:00

An Introduction to Bayesian Network Learning

Patrick Shaughnessy

Bayesian network learning is a relatively new machine learning technique. This talk will briefly introduce the topic of machine learning and then present an overview of Bayesian networks, a method for modeling dependencies. Finally, Bayesian network learning will be discussed. Bayesian networks are a graph data structure which model statistical relationships. Bayesian networks are useful for explaining and summarizing data to humans and for making predictions about unknowns in new observations. Research into Bayesian networks is ongoing, particularly in automating the construction of the topology of the graph using both data and prior knowledge.

Finding Difference Networks in Gene Expression Data

Thao Nguyen

Biological data from DNA microarray is growing dramatically; therefore, it is impractical to analyze biological data manually. Fortunately, machine learning offers the capability of automatically analyzing the huge amount of data that gene-expression microarray technology can produce. For this project, we have been using Bayesian networks, one of the learning systems provided by machine learning, to find the gene-gene interactions that differ in two populations. In this presentation, we outline research for identifying difference networks that is based upon Bayesian network and present some preliminary results taken from cancer gene-expression data.

Using Machine Learning to Improve the Identification of Genes

Chun-Yin (Cathy) Liu

The analysis of the human genome is of vital importance because it is now well understood that genes control cellular processes. One important task in analyzing genomes is the identification of genes. We discuss the use of machine learning to improve the gene identification tool GLIMMER, which, although it finds ~97% of all genes in a genome when compared with published annotation, has a high false positive rate, sometimes reporting as many non-genes as confirmed genes.

Our research plan, which has just begun, is to evaluate GLIMMER first-hand by applying it to the genome of an algae virus and evaluating its performance. Next, we plan to use machine learning methods to refine GLIMMER's predictions and to suggest improvements to GLIMMER.