

UW CENTER FOR PATTERN ANALYSIS AND MACHINE INTELLIGENCE

GRADUATE SEMINAR SERIES

Synthesizing and Analyzing Align Pattern Clusters

Speaker: En-Shiun Annie Lee

Date: December 16, 2013

Time: 4:00pm – 4:30 pm

Place: E5 (5128) Refreshments will be served

Abstract :

Research Problem: Protein sequences from a family were studied by synthesizing sequence patterns and analyzing the results. The following three research questions were asked: (1) What do patterns and their variations mean in terms of biological function (such as structure, evolution)? (2) If biological information about the sequence is available with varying degrees of accuracy, can they be used? (3) How can these patterns with variations be visualized, represented, and analyzed?

Methodology: To find the protein sequence pattern, a novel process was developed to: (1) find a reduced set of patterns; (2) group those patterns by how similar they are; (3) verify results by class labels, either explicitly or implicitly; and (4) analyze results by further organization.

Results: Our results show protein patterns and conserved amino acids are biologically functional and that variable amino acids have divergent functionality due to mutations. This divergence is proven by including biologically tagging each sequence by a function (such as gene function or species), whether it be objective or inherent. Most importantly, data represented by the pattern is crucial for considering the subset of sequences the pattern covers in order to obtain precise information theoretic measures and unsupervised classification.