

UW CENTER FOR PATTERN ANALYSIS AND MACHINE INTELLIGENCE

GRADUATE SEMINAR SERIES

Revealing Protein Classes by Amino Acid Variations in Sequence Patterns

Speaker: En-Shiun Annie Lee

Date: May 29, 2013

Time: 4:30pm – 5:00 pm

Place: E5 (5128) Refreshments will be served

Abstract :

Sequence patterns reveal important functionality of protein families and its amino acid variations identify its class functions. Supervised classification methods depend on class labels but cannot reveal embedded pattern variations related to inherent functionality. We use two measures, called sum of redundancy and class information gain that reveals sequence patterns and their amino acid variations without relying on the knowledge of the class labels beforehand.

We apply our unsupervised algorithm to the cytochrome complex protein and the spermidine/spermine-N1-acetyltransferase (SSAT) protein families and found that (1) Class entropy computes distribution amongst one class or all class, (2) Class information gain determines distinct amino acids associated with different classes, and (3) Sum of redundancy reveals important amino acid associations. Our algorithm finds amino acid variations and associations which are important for class-specific amino acid studies.