UW CENTER FOR
PATTERN ANALYSIS AND MACHINE INTELLIGENCE

GRADUATE SEMINAR SERIES

Comparing Clustering Algorithms for Grouping Aligned Pattern Clusters

Speaker: En-Shiun Annie Lee

Date: August 7, 2013

Time: 4:00pm – 4:30 pm

Place: E5 (5128) Refreshments will be served

Abstract :

Advances in high-throughput bioinformatics have provided a large influx of novel sequences, thus making the analysis of the sequences crucial. Proteins sequences are composed of amino acid alphabets; network clusters of protein regions can be analyzed to reveal inherent biological knowledge. The important protein regions are represented by frequent sequence patterns in protein families, which we represent with Aligned Pattern Clusters (APCs). When two conserved protein regions occur simultaneous in one protein, this implies that they interact within the protein. This co-occurrences is used to cluster APCs into APC Clusters. The purpose of this paper is to compare three clustering algorithms for finding these APC Clusters: 1) maximum spanning tree with minimal cut, 2) k-means clustering, and 3) spectral clustering. We compare the three clustering algorithms by performing three sets of experiments: synthetic dataset, two biological case studies, and a large-scale biological study. The biological results are further confirmed by their corresponding three- dimensional structure. Our method and results are currently being verified an experimental lab crystallizing an important proteins in immunology.

WATERLOO ENGINEERING

CPAMI