UW CENTER FOR
PATTERN ANALYSIS AND MACHINE INTELLIGENCE

GRADUATE SEMINAR SERIES

Co-occurring Patterns in Sequences from Pattern Clusters

Speaker: Sanderz Fung

Date: April 10, 2013

Time: 4:00pm – 4:30 pm

Place: E5 (5128) Refreshments will be served

Abstract :

Advances in massive throughput in bioinformatics have made discovering new protein interactions crucial thus resulting in a wide range of novel methods. We propose a new method for finding potential protein interactions within a protein based solely on finding the co-occurring patterns discovered from protein sequences from the same protein family. Our method first uses two established algorithms to find strong amino acid patterns and obtain Aligned Pattern Clusters (APCs) respectively. We then identify the patterns, from two or more APCs, which co-occur on a same sequence which we call pattern co-occurring sequences (PCS). We found and ranked APC pairs according to the number of PCS they share. Additionally, we created graphs based on the co-occurrence between the APCs to observe and find internal protein interaction pairs. We ran our method on three sets of protein family sequence data from the Pfam database and found one cluster pair in cytochrome c, one pair and one triplet in ubiquitin, and one triplet in triosephosphate isomerase that are most likely corresponding to internal protein interactions. The results are then validated in the 3D protein structure of the respective protein family. Surprisingly, when mapped back to the protein structure, we discovered that these clusters are close in spatial distance although they are not close to one another on the sequence. We conclude that our method is able to find potential internal interactions in the large amount of data generated by next-generation sequencing.

WATERLOO ENGINEERING

CPAMI