



SDSU
presents
a thesis defense for
Master of Science
degree in
Computer Science

Wednesday,
August 12, 2015
12:00pm
GMCS 418

Sivanarayana Gaddam

Selection of Glycan Markers for Cancer Detection Using Causal Feature Selection Algorithms

Abstract

Feature selection is critically important in bioinformatics applications where datasets have number of variables significantly larger than the number of observations. These areas include gene expression analysis, combinatorial chemistry, Proteomic Array analysis and Printed Glycan Array analysis. This thesis focusses on Printed Glycan Array datasets and attempts to implement a feature selection algorithms which provide an efficient and overfitting-free classification. Printed Glycan Arrays have revolutionized the analysis of the specificity of glycan binding proteins, providing information that simultaneously illuminates the biology mediated by them and decodes the information content of the glycome. Numerous research studies have shown that printed glycan arrays can be used as a tool to diagnose cancer and various viral diseases. The thesis proposes an implementation of the “Markov Blanket Discovery” which is specially optimized for the PGA datasets. The implementation is evaluated on real clinical data obtained from the School of Medicine of NYU, which contain 65 control samples of high-risk subjects exposed to asbestos and 50 subjects diagnosed with malignant mesothelioma. The results are compared to artificially generated data which have general characteristics similar to the original real data.

Thesis Committee

Marko Vuskovic, Thesis Chair, Department of Computer Science
Carl Eckberg, Department of Computer Science
Christopher Paolini, College of Engineering