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

Monday,
October 19, 2015

10:00am
GMCS 418

Ting Guo

Classification of Immunoprofiles
With Combined Correlated Features Algorithm

Abstract

In the recent decade, the development of bioinformatics algorithms for feature selection and classification based on the new biomarker platform called Printed Glycan Arrays (PGA) has become one of challenging researched topics. Since many of PGA signals are routinely highly correlated and the data matrices with a high column correlation can adversely affect the results, it is important to take into account this property of PGA chips. The goal of this thesis is to develop a new algorithm inspired by the ideas which followed the development of Correlation Adjusted Wilcoxon ranking (CAW) algorithm, called Combined Correlation Feature selection (CCF). The new algorithm combines correlated features in training set into new uncorrelated features which are subsequently used in feature selection and classification. The thesis presents the implementation of CCF algorithm and its analysis and testing on real data. As shown, the results of the presented experiments suggest a better performance of the CCF algorithm over CAW, and especially over straight-forward univariate feature selection without rejection of correlated features (SW), under certain properties of the correlation structure of data.

Thesis Committee

Marko Vuskovic, Thesis Chair, Department of Computer Science
Mahmoud Tarokh, Department of Computer Science
Barbara Bailey, Department of Mathematics & Statistics