Statistical Computing Methods for Imaging Data Processing

Organizers: *Don Hong* (dhong@mtsu.edu), Middle Tennessee State University, Murfreesboro, TN
John Wallin (john.wallin@mtsu.edu), Middle Tennessee State University, Murfreesboro, TN
Qinag Wu (qwu@mtsu.edu), Middle Tennessee State University, Murfreesboro, TN

Abstract

Many high dimensional data sets such as imaging mass spectrometry (IMS) and functional magnetic resonance imaging (fMRI) data are of the hyper-spectral imaging (HSI) type. Advanced mathematical tools and statistical techniques not only provide significance analysis of experimental data sets but also can help in finding new data features/patterns, guiding biological experiments designs, as well as leading computational tools development. In this mini-symposium, we would like to take this opportunity to exchange rearch ideas on statistical computing methods for imging data processing. Topics incude machine learning, multivariate statistical tools for imaging data analysis and feature selection or pattern recognition.

Session 1:

- **Speaker:** *Jingsai Liang and Lu Xiong* (jl4z@mtmail.mtsu.edu, lx2c@mtmail.mtsu.edu), Middle Tennessee State University, Murfreesboro, TN

  Title: IMSmining and Markov Random Field for Imaging Mass Spectrometry Data Analysis

  Abstract: In this presentation, we'll first introduce IMSmining, a newly developed software package by a research group at MTSU for imaging mass spectrometry processing using statistical methods. It contains functions of data visualization, biomarkers selection, and classification. Statistical algorithm selections include principle component analysis (PCA), support vector machine (SVM), LASSO, SPCA, elastic net (EN), and weighted EN (WEN). It can be used in either MATLAB GUI or MATLAB function interface. Then, we will talk about how to incorporate spatial information in IMS data analysis using Markov Random Field (MRF) and Bayesian method. Markov chain Monte Carlo (MCMC) is used to for computing implementation and maximum pseudo likelihood method is used for parameter estimation. This is joint work with Don Hong.

- **Speaker:** *Richard K. Archibald* (archibaldrk@ornl.gov), Oak Ridge National Laboratory, TN

  Title: Detection of Movement from Single Trial EEG

  Abstract: Single-trial analysis of EEG data has encouraged development of brain computer interfaces (BCI). Many available BCIs use the underlying functional and anatomical structure of motor cortical areas, but only grossly, as for example, by exploiting EEG frequency bands recorded separately from the two hemispheres to control devices. Here, we explored the predictive power of EEG in the context of leftward and rightward directed hand movements performed only with the right hand.

- **Speaker:** *Jianzhong Wang* (mth_jxw@shsu.edu), Sam Houston State University, Huntsville, TX
  Title: Data classification using wavelets on data trees

  Abstract: Let $X \subset \mathbb{R}^D$ be a set of high-dimensional data with the cardinality $|X| = n$, and $C = \{C_1, C_2, \cdots, C_s\}$ be the set of classes. A classifier is a function $f : X \to C$ that classifies all vectors in X into s different classes. Suppose that a subset $S \subset X$ is labeled by class, that is, f is known on the set S, but unknown on $X \setminus S$. The classification problem is to find f on the set X by the structure of X. In this presentation, we use a hierarchical data tree approach to the classifiers. First, a hierarchical data tree is constructed based on the weighted data graph. Then an orthogonal wavelet basis is constructed on the data tree so that each classifier f has a unique wavelet decomposition. Finally, an interpolation scheme is applied to find the classifier f on X from its values on the label set S.

- **Speaker:** *John Wallin* (John.Wallin@mtsu.edu), Department of Physics and Astronomy & Center for Computational Science, Middle Tennessee State University, Murfreesboro, TN

  Title: : Automatic Pattern Recognition and Citizen Science

  Abstract: Some of the most difficult task in science involve processing image data. In many cases, human experts still do better than computers on these advanced tasks. In this talk, I will discuss how crowd sourcing complex image tasks to Citizen Scientists is being used to classify and analyze scientific images and the algorithmic approaches needed to make crowd sourcing of complex image analysis tasks successful. I will also present examples where automatic image classifiers are working in conjunction with Citizen Science projects. In particular, I will show results from WNDCHRM, a multi-purpose image classification system that was developed by Dr. Lior Shamir. This code uses a brute force approach to image analysis and classification, using a combination of statistical methods for both supervised and unsupervised learning. The results from WNDCRHM will be compared to the results of from Citizen Science volunteers on several complex imaging processing tasks.

Session 2:

- **Speaker:** *Sayan Mukherjee* (sayanmuk@gmail.com), Duke University, Durham, NC

  Title: Supervised Dimension Reduction

  Abstract: We look at the problem of supervised dimension reduction (SDR) from three perspectives. The SDR problem is given observations, x-y pairs, infer a subspace of the input data without losing any predictive accuracy. We first study it as a geometric problem, show that the gradient of the regression or classification function is a central quantity, and in the case where the inputs are concentrated on manifold the covergence of estimators will be a function of the dimension of the manifold and not the ambient space. We then examine the same problem from a probabilistic modeling perspective and give a Bayesian solution. In this setting factor models and the Grassmann manifold are central. We close with an algorithmic perspective where we show that randomized algorithms can be used for SDR on massive data. We discuss the statistical implications of these randomized algorithms, they are a form or regularization. We illustrate the ideas with applications in imaging.

- **Speaker:** *Ying Guo and Ran Shi* (yguo2@emory.edu), Emory University, Atlanta, GA

  Title: A hierarchical group ICA regression model for fMRI data

  Abstract: Independent component analysis (ICA) has become an important tool for identifying and characterizing brain functional networks in functional magnetic resonance imaging (fMRI) studies. A key interest in such studies is to understand between-subject variability in the distributed patterns of the functional networks and its association with relevant subject characteristics. With current ICA methods, subjects covariate effects are mainly investigated in post-ICA secondary analyses but not taken into account in the ICA model itself. We propose a new hierarchical group ICA regression model that can directly model subjects covariate effects in decomposition of multi-subject fMRI data. Our model provides a formal statistical method to examine how spatial distributed patterns of functional networks vary among subjects with different demographical, biological and clinical characteristics. We propose a maximum likelihood estimation method for the new group ICA regression model.

- **Speaker:** *Feng Liang* (liangf@illinois.edu), UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, IL

  Title: fMRI analysis with structured Bayesian variable selection

  Abstract: Functional magnetic resonance imaging (fMRI), one of the recently developed forms of neuroimaging technology, has radically improved our understanding of the structure, functional organization and pathology of the human brain. One main aim of fMRI is to detect regions that respond to certain stimuli, i.e., activation regions. Besides the enormous data size and complex structure–both temporal and spatial–of fMRI data, one challenge in fMRI analysis is that the fMRI response is not instantaneous, but lagged and damped by an unknown hemodynamic response function (HRF). Although HRF is often assumed to be given by some Gamma approximation in previous fMR analysis, modeling HRFs in fMIR experiments is getting an increasing attention recently due to its importance in understanding neurophysiology. We present a novel Bayesian method for simultaneous HRF estimation and activation detection for fMRI data. A Bayesian variable selection approach is used to induce shrinkage and sparsity; a spatial prior on latent variables is used to pool information from neighboring voxels; the activation map is generated from the full spectrum of posterior inference through a Markov chain Monte Carlo scheme. By integrating functional activation detection and HRFs estimation in a unified framework, our method is more robust to noise and less sensitive to model mis-specification, as demonstrated by simulated and real examples.

This talk is based on joint work with J. Xia and Y. Wang from UIUC.

- **Speaker:** *Qiang Wu* (qiang.wu@mtsu.edu), Middle Tennessee State University, Murfreesboro, TN,

  Title: Empirical mode decomposition and its application in imaging processing

  Abstract: Empirical mode decomposition (EMD) can be used to produce multiscale decomposition of the data. It serves as an alternative to the more traditional Fourier series and wavelet analysis. EMD decomposes signals or images into a trend function plus a finite, often small, number of components called intrinsic mode functions. Roughly speaking, the trend function describes the contents while the higher order components capture the details. While EMD may have various potential uses in imaging processing, I will focus on its application in visual stylometry in this talk.