Editorial **Open Access**

Hyperspectral Imaging Type Medical Data Processing: Spectral and Spatial Analysis

Don Hong[1,2*], Zhihui Yang[2] and Jiancheng Zou[1,2]

[1]Center of Computational Sciences, Middle Tennessee State University, Murfreesboro, Tennessee, USA
[2]College of Sciences, North China University of Technology, Beijing, China

Over the past decade, hyperspectral imaging has quickly grown into a powerful and versatile technology. It generates images of narrow spectral bands over a continuous spectral range, and produces the spectra of all the pixels in the scene. Using hyperspectral imaging data allows for the comparison of spectral similarity and the extraction of sub-pixel scale information. Hyperspectral imaging technology, along with advanced mathematical, statistical, and computational techniques for data analysis, has been applied to many real world applications. Imaging mass spectrometry (IMS) and functional magnetic resonance imaging (fMRI) are forms of hyperspectral imaging with wide ranges of application in recent medical sciences. All these types of imaging data share common properties of high-dimensionality and complex spatial correlation structure. To fully exploit the advantages of these complex data sets, sophisticated methods of data processing and analysis must be developed. The generation and analysis of high-dimensional, spatially explicit, and spectral data presents a significant computational challenge. In particular, there is a major need for automated and fast information extraction algorithms that are reliable and consider both the spectral and spatial information. Advanced mathematical tools and statistical techniques can, not only assist in the analysis of experimental data sets but can also help in finding new data features/patterns, thereby guiding the design of biological experiments, and leading in the development of computational tools.

The development of mass spectrometry (MS), such as matrix-assisted laser desorption ionization (MALDI) time-of-flight (TOF) MS, and imaging mass spectrometry (IMS), greatly speeds up proteomics research. An image/spectral data cube is a three dimensional array in which the X and Y coordinates represent spatial (image) information and the Z coordinates contain spectral (mass over charge m/z) data. Peptides and spectral trends found in the data restricted mass spectrometry images can also be modeled in the m/z domain. Similarly, trends observed in m/z and resolution found during its limited observation may be modeled in the spatial domain. A combination of the three data sets may also yield new insights into the surface composition via spectral analysis. In order to make full use of IMS data, it is desirable to have global analysis of all these ion images. It is necessary to utilize both m/z resolution of the mass spectrum in every pixel and the spatial information for the whole MS image. This leads to many challenges and difficulties in MS and IMS data processing [1]. It is critical to find a sparse representation for the original data to reduce its high dimensionality [2]. Noise in the MS/IMS data presents another challenge [3,4]. Multi-resolution analysis techniques [5,6], Markov random fields, dimension reduction kernel and multi-task learning methods together with multivariate analysis techniques such as elastic net [6] are used for IMS data processing and feature selection.

The fMRI uses a MR scanner to measure the blood-oxygenation-level dependent (BOLD) signal that is correlated with neural activity in response to some input stimuli. Such scans produce a sequence of 3D images each of which has on the order of voxels and the sequence contains hundreds/thousands of time points. It provides not only a deep insight into the underlying dynamics of human brain, but also provides a basis for making inferences on brain activation regions. The goal of fMRI data analysis is to detect correlations between brain activation states and tasks that the subject performs during the scan. It also aims to discover correlations between specific cognitive states, such as memory and recognition, which are induced in the subject. One common statistical model used for fMRI data analysis is the generalized linear model (GLM) model, which analyzes each voxel's data separately [7]. However, neighboring voxels are likely to activate or deactivate together, partly because diffusive blood flow links proximate neurons, and partly because of various preprocessing procedures, such as smoothing, intentionally correlate neighboring voxels. Statistical models such as multivariate pattern analysis (MVPA) [8] and spatial Bayesian [9,10], take a different approach, utilizing relationships across voxels. These methods partition the data set of voxels according to experimental conditions. A training model that summarizes the information in the voxels is created to help discriminate between the conditions. The model is run on test data to link experimental conditions to particular brain activity patterns. Unlike principle component analysis (PCA) which identifies the directions of maximal variance, independent component analysis (ICA) identifies the directions that maximize independence. While ICA has been successfully applied to single-subject analysis, the extension of ICA to group inferences is not straightforward and remains an active topic of research. A unified framework for estimating and comparing group ICA models with varying spatiotemporal structures is proposed for fMRI study in [11]. Ignoring the spatial correlation may prevent inclusion of correlated variables in the classification model. To account for the special spatial structure in the fMRI data, a spatially varying coefficients model [12] can be incorporated with a Bayesian framework [10]. Sparse dictionary learning is a relatively new tool in signal processing and it has been applied in fMRI data analysis [13]. A wavelet-based Bayesian nonparametric regression model with Dirichlet process and Markov random field's priors for fMRI data is proposed in [14]. Modern techniques in big data analysis such as machine learning and statistical computing will definitely play a more important role in hyperspectral imaging type medical data processing including biomarker discovery and classification.

*Corresponding author: Don Hong, Department of Mathematical Sciences, Box 34 MTSU, Murfreesboro, Tennessee, USA, Tel: +1-615-904-8339; E-mail: dhong@mtsu.edu

References

1. Hong D, Qin SY, Zhang FQ (2010) Mathematical Tools and Statistical Techniques for Proteomic Data Mining. Int J Math & Computer Sci 5: 123-140.

2. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15: 3736-3745.

3. Chen S, Hong D, Shyr Y (2007) Wavelet-Based Procedures for Proteomic MS Data Processing. Comput Stat & Data Anal 52: 211-220.

4. Chen S, Li M, Hong D, Billheimer D, Li H, et al. (2009) A novel comprehensive wave-form MS data processing method. Bioinformatics 25: 808-814.

5. Xiong L, Hong D (2015) Multi-Resolution Analysis Method for IMS Proteomic Data Biomarker Selection and Classification. British J Math & Computer Sci 5: 65-81.

6. Hong D, Zhang FQ (2010) Weighted Elastic Net Model for Mass Spectrometry Imaging Processing. Math Model Nat Phenom 5: 115-133.

7. Friston K, Ashburner J, Kiebel S, Nichols T, Penny W (2007) Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, Elsevier, USA.

8. Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7: 523-534.

9. Smith M, Fahrmeir L (2007) Spatial Bayesian variable selection with application to functional magnetic resonance imaging. J Am Stat Assoc 102: 417-431.

10. Xia J, Liang F, Wang YM (2009) FMRI analysis through Bayesian variable selection with a spatial prior. IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI '09, Boston, MA, 714-717.

11. Guo Y, Tang L (2013) A hierarchical model for probabilistic independent component analysis of multi-subject fMRI studies. Biometrics 69: 970-981.

12. Fan J, Zhang W (2008) Statistical Methods with Varying Coefficient Models. Stat Interface 1: 179-195.

13. Lee K, Tak S, Ye JC (2011) A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion. IEEE Trans Med Imaging 30: 1076-1089.

14. Zhang L, Guindani M, Versace F, Vannucci M (2014) A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. Neuroimage 95: 162-175.