Managing Big Data from MRI: the Physicist's Perspective

Hae-Jeong Park, Ph.D.
Yonsei University College of Medicine, Departments of Nuclear Medicine, Radiology, and Psychiatry, BK21 PLUS Project for Medical Science, Seoul, KOREA.

Recent advances in MRI technology have released a flood of medical image data and facilitated unprecedented discovery science (Van Horn and Toga, 2014). As MRI-based neuroimaging data inherit properties of 'Big data' in terms of volume, velocity, variety, veracity and value (Demchenko et al., 2012), it receives growing attention with the anticipation of opening a new era of neuroinformatics. Although the idea of big data emerges with the sharing of data publicly or across multiple centers (Poldrack and Gorgolewski, 2014), managing big data within a single clinical institute becomes a big challenge, due to the rapid accumulation of new data and the ever-growing demands for appropriate analyses. The goal of managing big data is to uncover unknown or previously unquestioned information of brain disease using data from a big population, called data mining. For this purpose, the first step of data-mining neuroimaging data begins with automated image preprocessing such as artifact removal, spatial standardization, contrast standardization, image parcellation or extracting regions of interest. The success for big data mining, especially for data with enormous data dimensions (i.e., variables), mostly depends on selecting informative features that sufficiently characterize data sets. Based on the reduced feature sets, underlying architecture or information can be data-mined via data clustering, classification or model construction for prediction. All of these machine-learning techniques require model training steps in supervised, unsupervised or semi-supervised ways, depending on the question of interest. Regardless of the type of training, training should be operated online in order to keep pace with the continuously accumulating clinical big-data. Efficient reporting and visualization of the extracted multi-dimensional information, alongside integrating meta-data from electronic health records are also mandatory. Considering the maturity of the automated analysis technique, connectome (individualized maps of whole brain inter-regional connectivity based on MRI) (Park and Friston, 2013), can be the most representative target for clinical big data mining because of its importance as biomarkers for brain disease and its high capability for big data. This presentation will introduce big data mining methods for neuroimage data from clinical MRI, especially for structural and functional connectome, address challenges and limitations of data mining in a local center, and place emphasis on the importance of asking appropriate questions for the big data era, expressed in the idiom “Big data with Big idea”.

[REFERENCES]

Demchenko Y, Zhao Z, Grosso P, Wibisono A, de Laat C (2012) Addressing Big Data challenges for Scientific Data Infrastructure. In: IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom 2012), pp 614-617 Taipei, Taiwan: IEEE Computing Society.

Park HJ, Friston K (2013) Structural and functional brain networks: from connections to cognition. Science 342:1238411.

Poldrack RA, Gorgolewski KJ (2014) Making big data open: data sharing in neuroimaging. Nat Neurosci 17:1510-1517.

Van Horn JD, Toga AW (2014) Human neuroimaging as a "Big Data" science. Brain Imaging Behav 8:323-331.