Procedia Computer Science

Volume 53, 2015, Pages 48–55

2015 INNS Conference on Big Data

Connectomics to Semantomics: Addressing the Brain's Big Data Challenge*

Xerxes D. Arsiwalla[1], David Dalmazzo[1], Riccardo Zucca[1], Alberto Betella[1], Santiago Brandi[1], Enrique Martinez[1], Pedro Omedas[1], and Paul Verschure[1,2]

[1] Synthetic Perceptive Emotive and Cognitive Systems (SPECS) Lab, Center of Autonomous Systems and Neurorobotics, Universitat Pompeu Fabra, Barcelona, Spain.
x.d.arsiwalla@gmail.com
[2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Abstract
Can semantic corpora be coupled to dynamical simulations in such a way so as to extract new associations from the data that were hitherto unapparent? We attempt to do this within neuroscience as an application domain, by introducing the notion of the semantome and coupling it to the connectome of the human brain network. This is implemented using BrainX^3, a virtual reality simulation cum data mining platform that can be used for visualization, analysis and feature extraction of neuroscience data. We use this system to explore anatomical, functional and symptomatic semantics associated to simulated neuronal activity of a healthy brain, one with stroke and one perturbed by transcranial magnetic stimulation. In particular, we find that parietal and occipital lesions in stroke affect the visual processing pathway leading to symptoms such as visual neglect, depression and photo-sensitivity seizures. Integrating semantomics with connectomics thus generates hypotheses about symptoms, functions and brain activity that supplement existing tools for diagnosis of mental illness. Our results suggest a new approach to big data with potential applications to other domains.

Keywords: Brain Connectomics, Data Mining, Virtual Reality

1 Introduction

In attempts to stem the data deluge, traditional approaches to big data usually rely on massive computational resources. Even though this approach to data crunching may often be necessary, it may not suffice to outpace the exploding rate of data production. Take for instance, data generated within neuroscience. The adult human brain has on average 86 billion neurons with about 10^{14-15} synapses [9], exchanging information continuously, thus giving rise to cognition and conscious behavior. Neuronal signaling is not the only form of communication in the

*All authors contributed equally.

Selection and peer-review under responsibility of the Scientific Programme Committee of INNS-BigData2015

doi:10.1016/j.procs.2015.07.278

brain. Molecular signaling cascades due to protein-protein interactions and tissue-specific gene expression patterns are also known to play a role in brain function and disease (especially, those originating during development). Precision measurements and simulations of these multi-scale processes are needed to understand causal mechanisms and dynamics of disease. However, this approach generates overwhelming amounts of data in the order of terabytes or greater, which is already at the limits of current computational technology.

This calls for alternative approaches to big data. One such attempt within neuroscience is the BrainX^3 platform [1], [5], [2]. BrainX^3 is a large-scale simulation of the human brain with real-time interaction, rendered in 3D in a virtual reality environment. The idea is to combine pure machine computational power with human intuition for the exploration and analysis of complex dynamical networks. On this platform, we have reconstructed a large-scale 3D simulation of brain activity in virtual reality [2]. The simulation is grounded on human cerebral structural connectivity data obtained from diffusion spectrum imaging studies [8]. Based on this dataset we model neuronal dynamics to simulate large-scale activity over the cortex. Users can interact with BrainX^3 by perturbing brain regions with transient stimulations to observe reverberating network activity, simulate lesion dynamics or implement network analysis functions from a library of graph theoretic measures. Within the immersive mixed/virtual reality space of BrainX^3 users can explore and analyze dynamic activity patterns of brain networks, both at rest or during tasks, or they can use the system for discovering signaling pathways associated with brain function/dysfunction, or they could use it as a tool for virtual neurosurgery. The specific simulations discussed in [2] reveal the spatial distribution of activity in the attractor state, how the brain maintains a level of resilience to damage, effects of noise and physiological perturbations. Knowledge of brain activity in these varied states is clinically relevant for assessing levels of consciousness in patients with severe brain injury.

The aim of this paper is to extend the functionality of BrainX^3 to operate as a full-fledged hypotheses generator for large neuroscience datasets. As is often the case with complex data, one might not always have a specific hypothesis to start with. Instead, discovering meaningful patterns and associations within the data might be a necessary incubation step for formulating well-defined hypotheses that can then be tested with specific generative models. Hence, we introduce the notion of the *semantome*, a layer of semantics, conceived as a curated corpora of available information on brain regions, functions, diseases, etc. This layer is coupled to the connectomics of BrainX^3 in such a way that new associations in the semantomics layer are driven and shaped by neuronal activity generated in the connectomics layer. How does this work? Every individual brain region (nodes of the connectome) is well-documented in the literature and implicated in multiple cognitive functions and diseases. However, large-scale oscillations in the brain result from collective dynamics of locally coupled oscillators. Hence, stimulating a single sensory area activates related circuits that integrate sensory information. This serves as the basis of multi-modal integration in the brain. It is this integrated functionality that we seek to understand at the level of semantomics. To do this, we couple the anatomical network in BrainX^3 with available semantic corpora and develop a data mining engine to search across layers of information so as to generate meaningful associations between the neuronal simulation and semantic corpora. Based on the simulated activity, users can query the system for specific brain functions or diseases associated to activated pathways, while the data mining functionality searches and represents associated information either as text or graphs.