



Category: Bioinformatics

The unknown-unknowns: Revealing the hidden insights in massive biomedical data using combined artificial intelligence and knowledge networks

Chris Yoo*

Systems Imagination, Inc., Scottsdale, AZ, USA

Presenting and *Corresponding author: chris@siiventures.com

Abstract

Genomic data is estimated to be doubling every seven months with over 2 trillion bases from whole genome sequence studies deposited in Genbank in just the last 15 years alone. Recent advances in compute and storage have enabled the use of artificial intelligence techniques in areas such as feature recognition in digital pathology and chemical synthesis for drug development. To apply A.I. productively to multidimensional data such as cellular processes and their dysregulation, the data must be transformed into a structured format, using prior knowledge to create contextual relationships and hierarchies upon which computational analysis can be performed. Here we present the organization of complex data into hypergraphs that facilitate the application of A.I. We provide an example use case of a hypergraph containing hundreds of biological data values and the results of several classes of A.I. algorithms applied in a popular compute cloud. While multiple, biologically insightful correlations between disease states, behavior, and molecular features were identified, the insights of scientific import were revealed only when exploration of the data included visualization of subgraphs of represented knowledge. The results suggest that while machine learning can identify known correlations and suggest testable ones, the greater probability of discovering unexpected relationships between seemingly independent variables (unknown-unknowns) requires a context-aware system – hypergraphs that impart biological meaning in nodes and edges. We discuss the implications of a combined hypergraph-A.I. analysis approach to multidimensional data and the pre-processing requirements for such a system.

Citation: Yoo, C. The unknown-unknowns: Revealing the hidden insights in massive biomedical data using combined artificial intelligence and knowledge networks [Abstract]. In: Abstracts of the NGBT conference; Oct 02-04, 2017; Bhubaneswar, Odisha, India: Can J biotech, Volume 1, Special Issue (Supplement), Page 235. <https://doi.org/10.24870/cjb.2017-a220>