Discovery Informatics: Can AI Do Science?

Lawrence E Hunter
University of Colorado School of Medicine
Larry.Hunter@UCDenver.edu / @ProfLHunter

Abstract

Revolutions in the availability, scope and value of openly available scientific data have dramatically increased the potential for significant discovery by AI. Are we ready?

History

Artificial Intelligence has a long history of theorizing about scientific discovery processes, generally tested by experiments that attempt to recapitulate prior scientific discoveries (e.g. [Simon 1969]). More recently, application of technologies originally pioneered in artificial intelligence research, such as machine learning and natural language processing, have played an increasing role in novel scientific discovery. For example, post-genomic biomedical research is increasingly driven by informatics, and five of the ten most cited biomedical informaticians have AI backgrounds.[1] However, there are as yet no clear examples of novel scientific discoveries made by AI systems; perhaps the closest has been King's Robot Scientist, which embodies both reasoning and control of experimental equipment [King, et al., 2004; King 2010].

Opportunity

In the 2003 Engelmore Award lecture, I proposed that the creation and publication of a peer-reviewed scientific journal article be used as a test for computational intelligence. The opportunity for AI systems to make genuine scientific discoveries and to publish them in the same venues as human generated science have never been better.

Broad adoption of standards for scientific data access, interoperability and metadata have made vast stores of the raw materials for discovery available to computer scientists otherwise lacking laboratory skills or resources. This data is often highly valuable and incompletely analyzed. Important new discoveries have come from secondary analysis of publicly available information (e.g. [Pasquall, et al., 2013] in diabetes or [Jenouvrier, et al., 2009] in climate change).

Data sharing mandates, infrastructure and tools are increasingly available from government agencies and other sources, e.g. the US National Institutes of Health Big Data to Knowledge Initiative (http://www.bd2k.nih.gov), the European Union's ELIXIR (http://www.elixir-europe.org) in biomedicine and the World Bank's Climate Change Knowledge Portal (http://worldbank.org/climate-change). Major scientific publishers have issued calls for submissions based solely on reanalysis of publicly available datasets (e.g. *Nature Genetics* 2014). Novel philosophical theories of scientific discovery [e.g. Craver and Darden 2013] provide powerful new conceptual frameworks.

Discovery Informatics Workshop

The goal of this workshop is catalyze a pivot from evaluation of AI theories of scientific discovery via recapitulation to the autonomous automated generation of novel and significant scientific discoveries. Are we ready?

References

Craver, CF and Darden L. *In Search of Mechanisms: Discoveries Across the Life Sciences.* University of Chicago Press 2013.

Jenourvier, S. et al., 2009 Demographic models and IPCC climate projections predict the decline of the Emperor Penguin population. *PNAS* 106(6):1844-1847

King, R. 2004 Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427(6971):247-252

King R. 2010. Rise of the Robot Scientist. *Scientific American* 304(1):72-77

Nature Genetics 2014 Call for Data Analysis Papers 46:213

Pasquall, L. et al. 2014 Pancreatic islet enhancer clusters enriched in type 2 diabetes risk associated variants. *Nature Genetics* 46:136-143

Simon, HA. *The Sciences of the Artificial* MIT Press, Cambridge MA 1969.

[1] See http://allisonbmccoy.github.io/scholar-scraper/index-bmi.html