21st Century Science: The Age of Intelligent Algorithms?

Stephen Roberts

Machine Learning Research Group,
Oxford-Man Institute, University of Oxford,
Alan Turing Institute
& Mind Foundry Ltd.
The (big) data we generate

During 2008, the number of things connected to the Internet exceeded the number of people on earth. By 2020 there will be 50 billion.

Source: Mario Morales, IDC
The scale of data

All printed material in the world: 200 petabytes (2 x 10^{17} bytes)

All words ever spoken by human beings: 5 exabytes (5 x 10^{18} bytes)

Internet Traffic in 2015: 960 exabytes (1 x 10^{21} bytes)

[UC Berkeley School of Information, 2003; Cisco Visual Networking Index, 2011]
Big science, big data

Square Kilometer Array
> TeraByte / second

Grid of > 22,000 seismic stations – helping understand Earth’s interior dynamics

LSST > 30 TeraBytes / night

Vast array of detectors at CERN’s ATLAS experiment > TeraByte / sec
The driving force behind AI/ML
Data analytics as the 4th Paradigm

Science Paradigms

- Thousand years ago: science was empirical describing natural phenomena
- Last few hundred years: theoretical branch using models, generalizations
- Last few decades: a computational branch simulating complex phenomena
- Today: data exploration (eScience) unify theory, experiment, and simulation
  - Data captured by instruments or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files using data management and statistics
AI/ML in Science

1. Detection & discovery

2. Principles from data

\[ G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \]

3. Human-in-the-loop

4. Smart experiments
Kepler space telescope

(2009 – 2014)
EVIDENCE OF AN ASTEROID ENCOUNTERING A PULSAR

P. R. Brook\textsuperscript{1,5}, A. Karastergiou\textsuperscript{1}, S. Buchner\textsuperscript{2,6}, S. J. Roberts\textsuperscript{3}, M. J. Keith\textsuperscript{4,7}, S. Johnston\textsuperscript{4}, and R. M. Shannon\textsuperscript{4}

\textsuperscript{1} Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK; paul.brook@astro.ox.ac.uk
\textsuperscript{2} Hartbeesthoek Radio Astronomy Observatory, P.O. Box 443, Krugersdorp 1740, South Africa
\textsuperscript{3} Information Engineering, University of Oxford, Parks Road, Oxford OX1 3PJ, UK
\textsuperscript{4} CSIRO Astronomy and Space Science, Australia Telescope National Facility, P.O. Box 76, Epping, NSW 1710, Australia

Received 2013 October 20; accepted 2013 November 13; published 2013 December 19
Extracting Physics from data

Conservation laws & constraints can be inferred from data

\[ F = m \ddot{x} \]

\[
\frac{1}{c^2} \frac{\partial}{\partial t} E - \frac{r}{c^2} \frac{\partial}{\partial t} E = 0
\]

\[
\frac{1}{c^2} \frac{\partial}{\partial t} B - \frac{r}{c^2} \frac{\partial}{\partial t} B = 0
\]
Human-algorithm symbiosis
Big data: Square Kilometer Array, 10 petabytes compressed images/day

Noisy reports: Twitter, Typhoon Haiyan
Designing optimal experiments

Current Experiments → Posterior Model → Select Experiment(s) → Run Experiment(s)

→
So – where are we going?
Data Analytics

Beyond-human-scale knowledge extraction

physics, medicine & healthcare, biology, finance & economics, energy, social science, customer-facing e-commerce, digital marketing
Data Analytics

Computation, statistics and inference on massive scales

high performance computing at scale, statistics & inference, intractable likelihoods, huge inverse problems under uncertainty
Intelligent systems

New methods of computing under uncertainty
probabilistic programming, probabilistic numerics, scalable intelligent optimisation, automated algorithm creation
Intelligent systems

Strength in depth - vision, speech and beyond: understanding AI

deep understanding of deep learning, vision & beyond vision, cognitive systems with understanding of the world

Interpreting scenes find what and where
Intelligent systems

Autonomy at scale, intelligent sensing and situation awareness

unifying heterogeneous data at scale, ultra-large machine sensing, autonomous systems, deciding and acting in a complex uncertain world, human-in-the-loop systems
Complex Systems

Cyber-physical systems at IoT scale, quantum computation in the wild

understanding and usage of beyond-human-scale dynamic networks, realising the potential of massive-scale connected data, new ways of discovering meaning, topology & homology, understanding multi-scale structure
Complex Systems

Computational mathematics

numerical analysis & massive parallelism, stochastic analysis for multi-scale feature learning, rough-paths for sequential and multi-d data, manifolds, bifurcation structure and non-linear systems, topology for structure learning
Data Systems

Cyber-security at IoT scale, databases, ontologies and algorithms
provenance, data curation and understanding scalable ontologies,
databases for global-scale computing, algorithms and software verification
Quantum Computation

A revolution in progress?

BRUTE-FORCE SOLUTION: \( O(n!) \)

DYNAMIC PROGRAMMING ALGORITHMS: \( O(n^22^n) \)

SELLING ON EBAY: \( O(1) \)

Still working on your route?

Shut the hell up.
The Challenges

**Scalability** features large across domains, from inference to algorithms, from stochastics to numerics, in application from astronomy to zoology, from e-health to finance, social data & e-commerce

Creating and **verifying** systems that **decide** and **act autonomously**, intelligently fusing **heterogeneous** data and coping with **scale**, **complexity** and **uncertainty**.

We have the opportunity to catalyse new ways of computing, from **quantum** systems to **probabilistic programming**.

We must be thought leaders in the **ethics** of information and data analytics – **ethics** must be a centrally embedded research stream – not a service or afterthought.
Automation is happening now...

The Economist
Coming to an office near you...

What today's technology will do to tomorrow's jobs

- Management, Business, and Financial
- Computer, Engineering, and Science
- Education, Legal, Community Service, Arts, and Media
- Healthcare Practitioners and Technical
- Service
- Sales and Related
- Office and Administrative Support
- Farming, Fishing, and Forestry
- Construction and Extraction
- Installation, Maintenance, and Repair
- Production
- Transportation and Material Moving

Graph showing employment against probability of computerisation.
The future of science?

Automated systems can review data, extract meaning & find explanations faster & at scales that we can only dream of.
The era of human science has ended.

The era of the Automated Scientist has begun.

With particular thanks to

Mike Osborne, Suzanne Aigrain, Aris Karastergiou, Chris Lintott, Matt Jarvis, Edwin Simpson, Steve Reece

and, of course, the algorithm...