Flipping the Light Switch

Using the Cambrian Explosion as a Metaphor for Accelerated Scientific Discovery

Gully A. Burns

Intelligent Systems Division, Information Sciences Institute
Paleontological Treasure Troves

Locations of Main Burgess Shale-Type Deposits

- Burgess Shale
- Sirius Passet
- Utah
- Kaili
- Chengjiang
- Emu Bay Shale

© Royal Ontario Museum

(from http://burgess-shale.rom.on.ca/)
The Cambrian Explosion

Every metazoan phylum came into existence during this period.
Andrew Parker hypothesizes that the cause of the dramatic evolutionary radiation was the evolution of the early eye and visual system. This made predatory behavior much more successful, leading to defensive specializations and transformation from precambrian forms (without skeletons) to armored and more highly evolved Cambrian animals (skeletons, spikes, teeth, swimming capabilities and armor).

Using the ‘Light Switch’ Metaphor for Scientific Discovery

Consistent with this hypothesis, we might attempt to look for:

• ‘Light’ – a ubiquitous data signal with rich substructure
• ‘Evolution’ – Accelerated scientific discovery
• ‘Eyes’ – Experimental methods + data gathering
• ‘Visual System’ - Data analysis and statistics
• ‘Intelligence’ – Scientific Theory
‘Cancer in the Age of Algorithms’
– a success story

Dr. Shirley Pepke had late-stage ovarian cancer and used advanced unsupervised learning to help her fight her cancer (with a colleague at ISI: Dr. Greg ver Steeg)

“It was suddenly like looking at the dictionary of tumor biology. It was suddenly pulling out all this information other algorithms couldn’t. It looked really beautiful and informative.”

Dr. Pepke switched treatments and her cancer is in remission

Dr. Shirley Pepke and Dr. Greg Ver Steeg

Visual System: Greg ver Steeg’s ‘Correlation Explanation’ (CorEx)

This factor “explains” the relationships in the variables it is connected to.

Search for the “most informative” factors for explaining relationships in data.

Measurement Data Variables

https://arxiv.org/abs/1406.1222,
Group of genes associated with a latent factor

Expression levels for each patient (sorted by latent factor)

A fraction of the CorEX hierarchical structure
Gene Ontology Annotations show that groups reflect strong, specific, diverse functions.

- Positive regulation of cytokine production
- Leukocyte activation
- Regulation of immune response, blood coagulation
- Positive regulation of immune process; cell activation
- Leukocyte migration and activation
- Regulation of immune response
- Inflammatory response
- Leukocyte migration
- Regulation of lymphocyte proliferation
Where’s the Light Switch?

• ‘Light’ => Gene Expression
• ‘Evolution’ => Precision Medicine
• ‘Eyes’ => Gene Chip Cancer Assays
• ‘Visual System’ => CorEx (Unsupervised Learning)
• ‘Intelligence’ => Linkage to Gene Ontology (requires human curated models + interpretation)

How does this idea generalize in the broader context of accelerating science?
Scientific Discovery: From ‘Elements’ to the Elements

Observe phenomena (‘fire’)
Get data to characterize phenomena (what happens when you burn things?)
Postulate mechanism (phlogiston)
Find correct mechanism (oxygen + combustion)
Formulate theory (chemistry)
Export reasoning (periodic table)
Technology (e.g., combustion engine)

‘stuck’ paradigm
Cycles of Investigation ('KQED' Model)
The Complexity of Data Spaces

Neuroscience is an example of a complex multi-level domain

A variety of variables are needed to describe models across scales and systems.

Complex experiments have to be designed and executed by human experts to investigate and test these models.

How to apply the Cambrian Metaphor here?
Knowledge Engineering from Experimental Design

Gene Expression Lab Study
Khan et al. (2007)

Russ et al (2011), BMC Bioinformatics
E-Science Workflows

An example of a workflow for genomic analysis from the WINGS workflow system

http://www.wings-workflows.org/
Measurements and Metadata
Define the Data Space

Could we apply tools like CorEx to examine the structure of this space?

Would this be like suddenly flipping the switch and turning on the lights?
Acknowledgements

Greg ver Steeg
Oren Etzioni
Alan Watts
Prem Natarajan
Yolanda Gil
Ed Hovy

NIH, DARPA,
NSF, IARPA