Probabilistic Modeling Lab, Interdisciplinary Affective Science Lab
Northeastern University

Neuroscience, Sloppiness, and Ground Truth

(Or: could a computer scientist "Game Shark" a brain?)

Eli Sennesh
sennesh.e@husky.neu.edu

December 21, 2018

Outline

Motivation

Neuroscience with a known ground truth
- Engineered ground truth
- Methods and assumptions

Results: negative and positive
- Negative
- Positive
- What was not tried?

What do we try next?

Discussion

We would like to understand how the nervous system works, especially the brain. However, it “works” at many different scales:

- Individual neurons and glial cells (neurobiology)
- Cortical and subcortical circuits (computational neuroscience)
- Whole-brain networks (systems, theoretical neuroscience)
- “Hearts and minds” (affective/cognitive science/psychology)

Motivation
Scale and Sloppiness

Many models of complex systems exhibit *sloppiness*: only a few parameters determine almost everything about the data. Sloppiness and stiffness can supposedly enable “emergence”, effective theories, and reductionism.

But how can we know when we are investigating the few *stiff* parameters?

Motivation
Scale and Sloppiness

Ideal gases:

- Positions and momenta of $N \times 10^6$ molecules (Sloppy!)
- Ideal gas law and temperature (Stiff!)

Brains:

- Molecules (Sloppy!)
- Cells (Sloppy!)
- Circuits?
- Networks?
- “Hearts and minds”?
- Person and society?

Which scales give us stiff models?

Motivation
Sloppiness and Ground Truth

The best way is to *already know* the ground truth about the system at multiple scales. Then we can check how well we're doing at finding stiff models at measurable scales.

An engineered ground truth

Could a neuroscientist understand a microprocessor?

"Microprocessors are among those artificial information processing systems that are both complex and that we understand at all levels, from the overall logical flow, via logical gates, to the dynamics of transistors."

Spoilers: no

"We show that the approaches reveal interesting structure in the data but do *not* meaningfully describe the hierarchy of information processing in the microprocessor."

An engineered “organism” and its “behaviors”

Model organism

MOS 6502 microprocessor from the Atari Video Game System

Model behaviors/tasks

- *Donkey Kong* (1980)
- *Space Invaders* (1978)
- *Pitfall* (1981)

Only the bootloader was simulated.

Experimental methods for an engineered model organism

Experimental methods used:

- Connectomics
- Lesion studies on single transistors
- Tuning selectivity of transistors
- Spike word correlational structure
- Functional imaging and local field potentials
- Granger causality for functional connectivity
- Dimensionality reduction (factor analysis)

Most of these methods do not make strong theoretical assumptions. However, a wide variety of theories have been ostensibly supported *using* these methods.

What would it mean to understand?

Computing and Engineering

If we understand a video-game system, we can write a good "high-level emulator" for it. This has been done for systems like the Atari, as well as many/most other early game systems.

Philosophy of Mind

Andy Clark called this a *reconstructive* model: one that lets you throw away the original system and rebuild it from the model. A useful backport to philosophy of science?

Negative results

Most of them, unfortunately.

- Connectomics: more than one kind of transistor.
- Lesioning: some transistors were essential for some games.
- Tuning properties: "Even if brain areas are grouped by function, examining the individual units within may not allow for conclusive insight into the nature of computation."
- Local field potentials: "[I]t is very hard to attribute (self-organized) criticality to the processor."

Positive results

There were some.

- Granger causality: "[R]egisters really affect the accumulator and decoders really affect the status bits."
- Linear factorization: "[S]ome components relate to both the onset and offset (rise and fall) of the clock signal."

What was not tried?

- Psychophysics
- Electrophysiology
- Transcranial stimulation

Analogy to a computer chip

We hack games with a Gameshark: rapidly alternate between observing behavior, inspecting activity, and intervening to test ideas. What could we (*ethically*) do with the brain?

Testing theories vs finding patterns

A lot of these methods for studying the brain looked “model-free”. Instead of starting with a theory telling us what has to happen, they just try to detect patterns in what does happen.

This tells us regrettably little about the brain.

Methodological issue:

Data does not substitute for prior knowledge.

What if we have a surfeit?

What is our prior knowledge?

The *predictive mind* paradigm starts from powerful theoretical arguments to constrain many levels.

- Good prediction (low KL divergence) minimizes waste heat,
- An organism must continually regulate its body to maintain physical and genetic integrity against entropy,
- "Every Good Regulator of a System Must Be a Model of That System",
- *Therefore*, an organism must predictively regulate its body (allostasis),
- *Therefore*, most organisms with nervous systems will have something like affect.

Sloppy models for complex systems

- The brain-body/heart-mind system appears sloppy at many scales.
- In sloppy model hierarchies, higher-level models describe the stiff parameters of lower-level ones.
- So wherever we know something “stiff”, at any level, we want it to constrain our models at all levels.
- Idea: model the hierarchical connections between models.

How can we use sophisticated prior knowledge to more tightly constrain our experimental models?

Example question

Are all the bottom-up signal gain levels in the brain *really* just Gaussian precisions?

Questions, comments, discussion

Let’s talk!

References

- “Could a Neuroscientist Understand a Microprocessor?”
- “Parameter Space Compression Underlies Emergent Theories and Predictive Models”
- “Radical Predictive Processing”
- “Every Good Regulator of a System Must Be a Model of That System”
- “The Thermodynamics of Prediction”
- “Allostasis: a model of predictive regulation”
- “Evidence for a large-scale brain system supporting allostasis and interoception in humans”
- How do Gameshark codes work?