BS-CS double major seminar

2024, Lecture 3:

Data science applications in Neuroscience

Yocheved Loewenstern

Room: 418 Email: yloewenstern@gmail.com

Outline

- Data
- Data science
- Data science in neuroscience

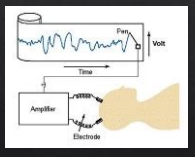
What is **data?**

- **Data** is factual information used as a basis for reasoning, discussion, or calculation which typically includes useful and irrelevant or redundant information and must be processed to be meaningful.
- **Information** is the state of the system.
- **Data analysis** is the act of transforming data with the aim of extracting useful information and facilitating conclusions.

(Adapted from: Merriam-Webster www.m-w.com & Wikipedia www.wikipedia.com)

Data analysis example

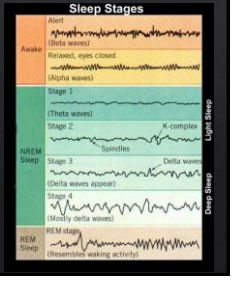
- Data: EEG (electroencephalogram)



- Data processing: Amplification, filtration, sampling & quantification.

Data analysis example

- Information: The state (sleep-wise) of the patient in one of the known stages.
- Data analysis: Based on assessing the power, pattern and frequency of the EEG and fitting to known sleep stages the state of the patient is found.

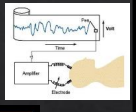


Data vs. Information

- Data: Electroencephalogram (EEG)

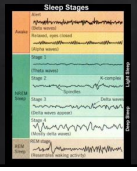
2 bytes/sample * 1000 samples/s * 60 seconds

= 120KB ≈ 1 Mbit



- Information: Sleep states

7 states ≈ 3 bits



This is not **ALL** the information in the data, but it is the **relevant** information.

Data analysis is **not** new

- Observing data forms the basis of science and inventions.
- Quantitative approaches to data analysis developed rapidly during the 20th century.
- Many of the major issues related to data analysis are not new either…

Ronald Fisher
The father of modern statistics

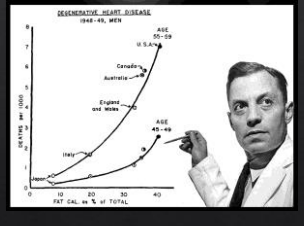
1935 The design of experiments

"Correlation does not imply causation"

Objectivity in data science:
Avid smoker who fought studies linking smoking and cancer

Ethics in data science:
1. Supporter of eugenics
2. Maintained that races are not equal intellectually

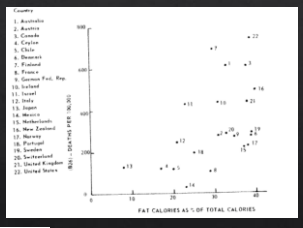
Data is useful for drawing conclusions



(Ancel Keys, 1953)

13,000 subjects total, 5-40 years follow-up.

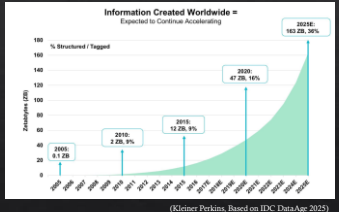
However, data does not necessarily solve controversies



Fat in the Diet and Mortality from Heart Disease

Big Data

- Explosion in the amount of data
- Sensors & devices
- Social media
- Enterprise data



(Kleiner Perkins, Based on IDC DataAge 2025)

KB	MB	GB	TB	PB	EB	ZB
10^3	10^6	10^9	10^{12}	10^{15}	10^{18}	10^{21}

Big data properties

The major change in data properties:

- Volume – number of data elements and datasets
- Velocity – speed of new data generation and moving around
- Variety – multiple types of data which are mostly unstructured
- Veracity – increase in the messiness and decrease in trustworthiness of data

(Bernard Marr – Big Data)

Outline

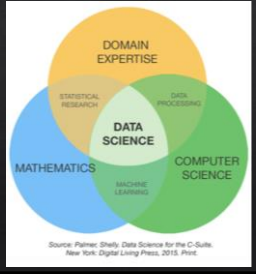
- Data
- Data science
- Data science in neuroscience

What is **data science** ?

- "**Data Science** is a new term. But in the same sense as Columbus was discovered NEW continent 1000 years ago" (Hector Garcia-Molina, Stanford)
- Extremely ill-defined field…
- Data science - Management, manipulation, extraction, and interpretation of knowledge from large amounts of data
- Data science is a multidisciplinary field of study stemming from statistics, computer science, engineering and other disciplines aiming at addressing the challenges of big data

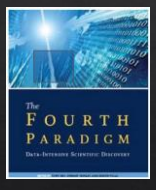
Harvard Business Review (2012) - Data scientist - the sexiest job of the 21st century

Data science – interaction of multiple fields



The fourth paradigm (Tony Hely 2009)

- Thousand years ago – **Experimental Science**
  - Description of natural phenomena
- Last few hundred years – **Theoretical Science**
  - Newton's Laws, Maxwell's Equations…
- Last few decades – **Computational Science**
  - Simulation of complex phenomena
- Today – **Data-Intensive Science**
  - Scientists overwhelmed with data sets from many different sources



Data-science ←→ Machine learning

Same or different???

Outline

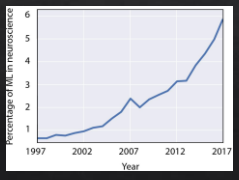
- Data
- Data science
- Data science in neuroscience

Data in Neuroscience

- Data in neuroscience is ill defined as:
  "anything which has to do with brain and/or behavior".
  - Physiology
  - Anatomy
  - Biochemistry
  - Psychophysics
  - Psychology
  - Ethology
  - …

Data science in Neuroscience

- Usage of data science and especially machine learning is growing in all fields of neuroscience (and especially systems neuroscience).

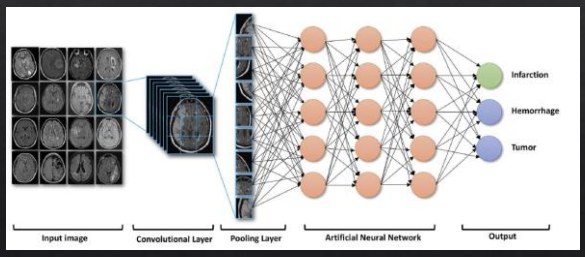


(Glaser et al., ProgNeuro, 2019)

Data science
Structural neuroscience (Neuroanatomy/Neurohistology)

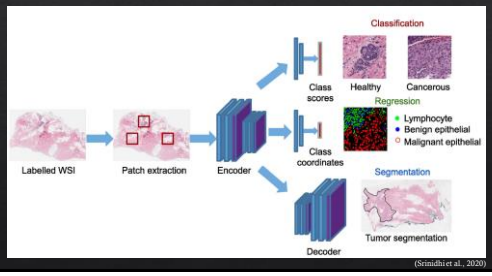
- Clinical science
  - Classify pathological states
  - Segment pathological areas
  - …
- Basic science
  - Identify cell types and sub-cellular components
  - Reconstructing morphology
  - …

Classification of different states based on MRI images



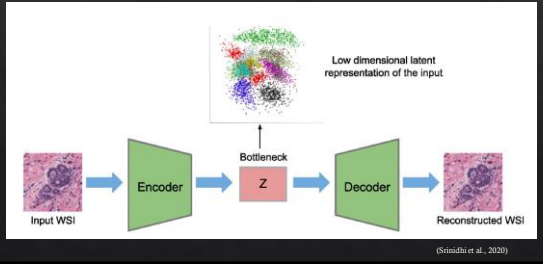
(Zaharchuk et al., 2018)

Supervised learning for histology



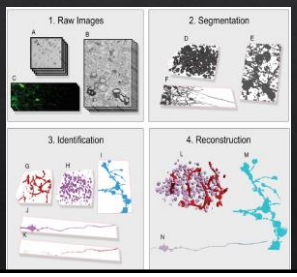
(Srinidhi et al., 2020)

Unsupervised learning for histology

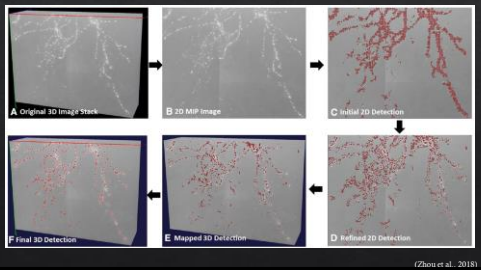


(Srinidhi et al., 2020)

Neuron reconstruction



DeepNeuron - Neuron tracing

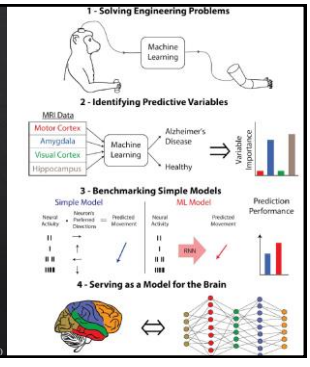


(Zhou et al., 2018)

Supervised learning in systems neuroscience

1. Engineering– using neural signals for interaction with computers/machines.
2. Identify brain areas (and/or neurons) predictive of a classification.
3. Relating simple models (population code or simple MAP model) to ML.
4. Comparison of brain implementation to ML implementation.

(Glaser et al., ProgNeuro, 2019)



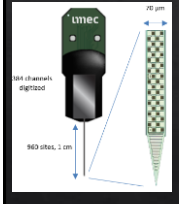
Neuroscience data
Electrophysiology

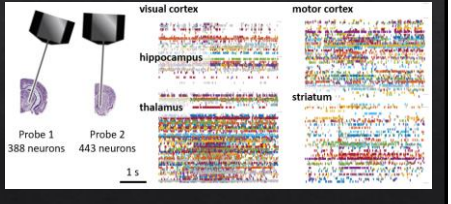
- Historically, a few neurons were recorded simultaneously for short periods.
- 25 years ago: ~100 electrodes → dozens of concurrently recorded neurons
- Currently: ~1000 electrodes → hundreds of concurrently recorded neurons for prolonged periods.

- Raw data – a typical experiment of 50 one-hour sessions:

960 contacts * 40K samples/s * 2 bytes/sample * 3600 seconds * 50 sessions =
13,824,000,000,000 → 13.8 TB

Neuroscience data – Electrophysiology
Neuropixels

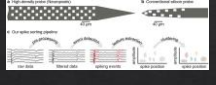




UCL, Neuropixels example data

Neuroscience data – Electrophysiology
Current methods do not scale up

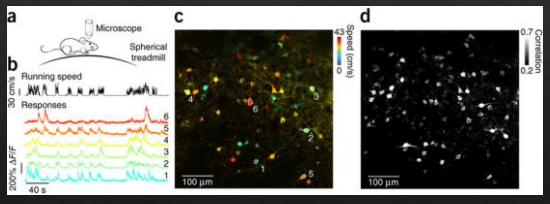
- Data pre-processing, e.g. spike sorting



- Data presentation: manual examination fails for large datasets.

- The complexity of neural interaction identification grows exponentially.

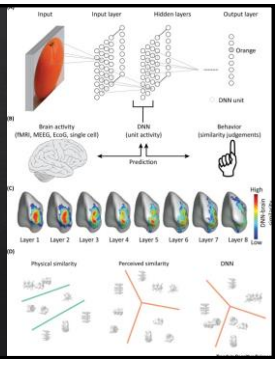
Neuroscience data
2 photon calcium imaging



Mapping brain activity at scale with cluster computing, Freeman et al. 2014, Nature Methods

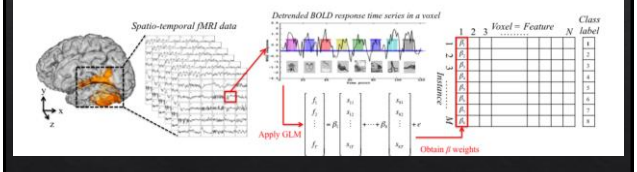
Decoding brain activity
DNN

Classification of brain activity to different sensory stimuli



(Cichy & Kaiser, TICS, 2019)

Decoding brain activity
Generalized linear model (GLM)

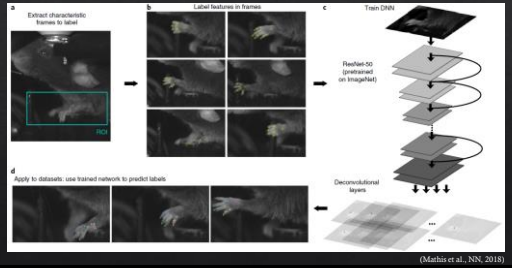


Voxel Selection Framework in Multi-Voxel Pattern Analysis of fMRI Data for Prediction of Neural Response to Visual Stimuli, Chou et al., 2014

Behavioral data

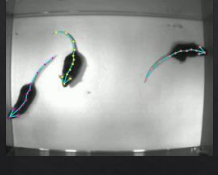
- Structured behavior
  - Limited number of possible behaviors
  - Highly constrained
  - Unnatural
- Unstructured behavior
  - Very large number of possible behaviors
  - Unstructured
  - Natural

DeepLabCut
Transfer earning from classic nets

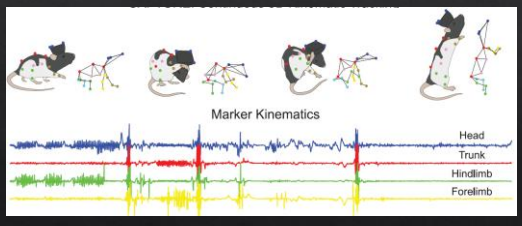


(Mathis et al., NN, 2018)

DeepLabCut

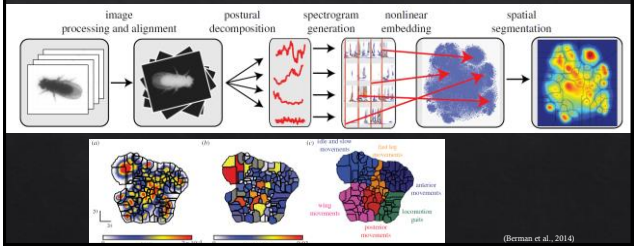


Neuroscience data
Behavior



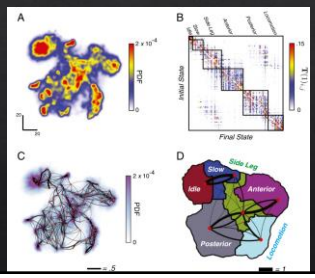
Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire, Marshall et al., 2021, Neuron

From pose to simple behavior



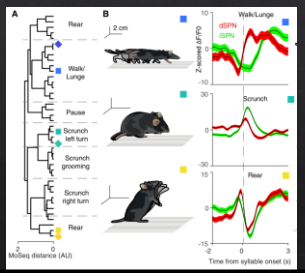
(Berman et al., 2014)

From simple to complex behavior



(Berman et al., 2016)

Linking behavior to neuronal activity

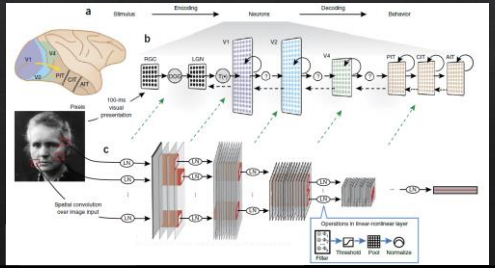


(Markowitz et al., Cell, 2018)

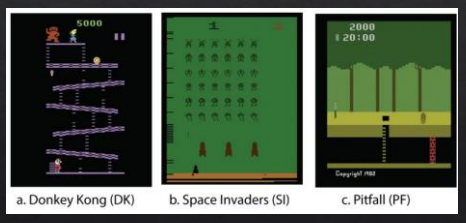
The reverse pathway

Can we use our understanding of computer science or data science to shed light upon neuroscience?

Relating DNN activity to the visual system



Could a Neuroscientist Understand a Microprocessor?



a. Donkey Kong (DK) b. Space Invaders (SI) c. Pitfall (PF)

(Joans & Kording, PLoS CompBiol, 2017)

Anatomy

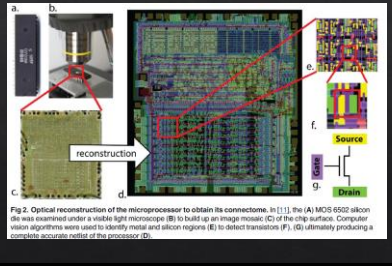
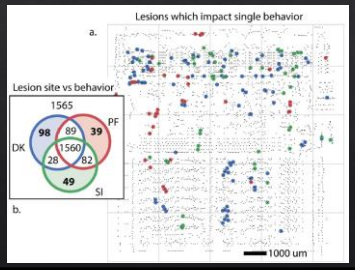


Fig 2. Optical reconstruction of the microprocessor to obtain its connectome. In [11], the (**A**) MOS 6502 silicon die was examined under a visible light microscope (**B**) to build up an image mosaic (**C**) of the chip surface. Computer vision algorithms were used to identify metal and silicon regions (**E**) to detect transistors (**F**). (**G**) ultimately producing a complete accurate netlist of the processor (**D**).

Lesions



Tuning curves

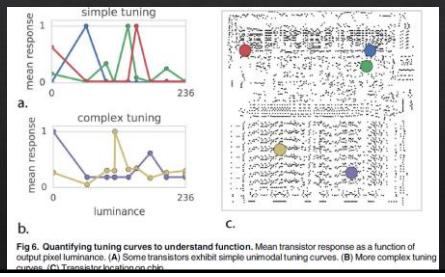
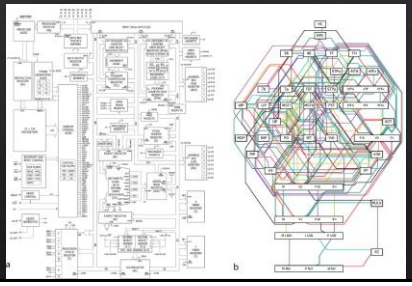


Fig 6. Quantifying tuning curves to understand function. Mean transistor response as a function of output pixel luminance. (**A**) Some transistors exhibit simple unimodal tuning curves. (**B**) More complex tuning curves. (**C**) Transistor location on chip.

Known functionality



Summary

- Extracting information from data is a major challenge in research.
- The exponential growth of the volume of unstructured data led to a focus on data science.
- Multiple fields within neuroscience provide huge datasets.
- Neuroscience utilizes key tools developed in other discipline to accommodate its datasets.
- Data science solutions may provide insights regarding brain data processing.