Neurons in the brain – illustration
Image credit: Benedict Campbell. Welcome Images

Connectomics!

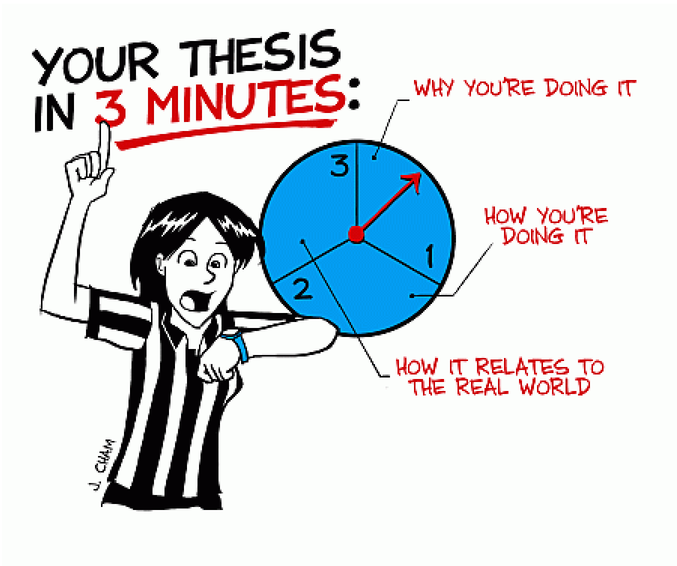
Neural Structure Reconstruction from Fluorescence Imaging of Neural Activity

Karan Singh

Pankaj Gupta

Objectives

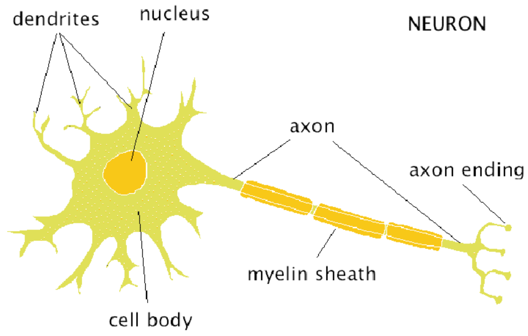
- **Communicate** research ideas and results.
- Study the **broad** perspective.
- You know what this is **about**!
- Keep the talk **enjoyable**.



Credits: **PHD** Comics

Some Biology?

- The **dendrites** receive signals from other neurons.
- A **synapse** allows signal transmission from one neuron to the target cell.
- **Neuro-transmitters** transmit signals across synaptic cleft.
- **Calcium** influx causes the release of neural transmitters.
- **Fluorescence** dye labelled Calcium allows study of neural activity.



Structure of Neuron
Image from Dr. C. George Boeree.

The Task

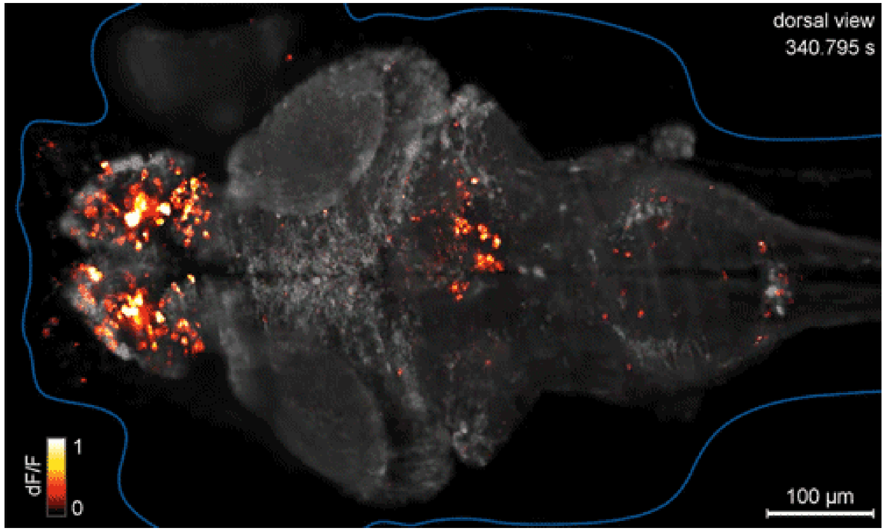
Data →

Time Series Activity of 1000 neurons

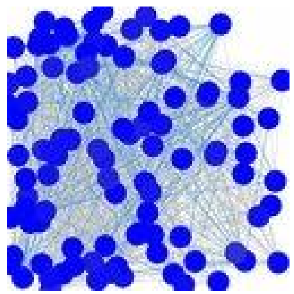
?

Predict →

Directed connections between neurons

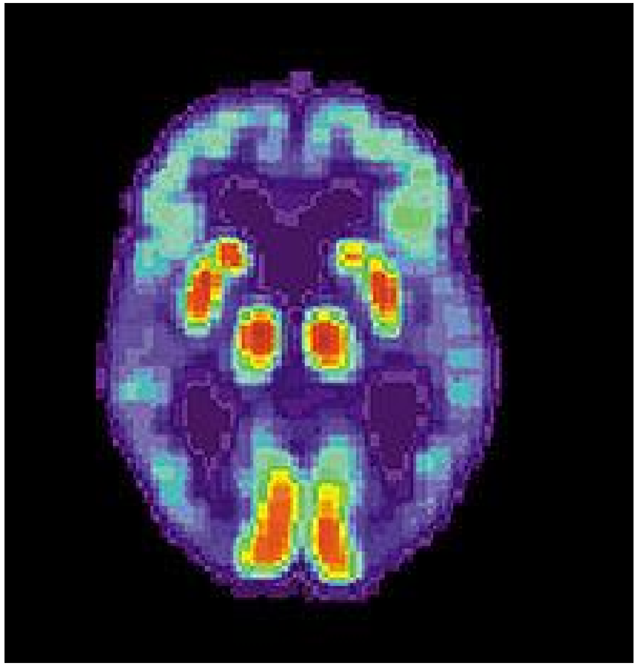


Brain of the zebrafish in action.
Credits: [Arens et al. Nature 485, 471–477 (May 2012).]



Why?

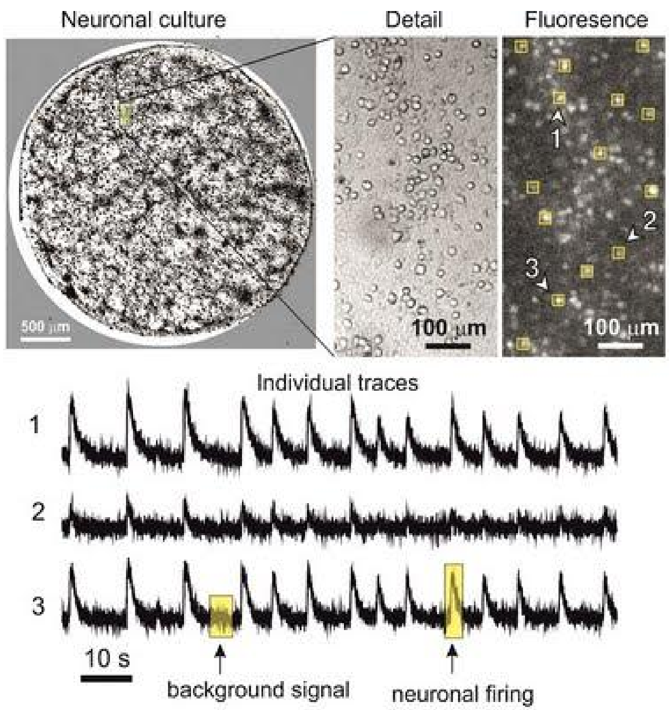
- Gaining a better understanding of the brain.
    - How does the brain learn?
    - Increase understanding of intelligent systems in general.
- Alterations in brain structure caused by diseases.
    - Treatment through a better understanding.
    - Epilepsy, Alzheimer's disease, Dementia.



PET scan of the brain of a person with AD
showing a loss of function in the temporal lobe
Credits: Wikipedia

Some Statistics

- Typical number of neurons ~ 100 billion
- Average number of synaptic connections ~ 7000
- Neuroanatomic methods of axonal tracing do **NOT** scale.
- Electron microscopy
    - Labour intensive
    - Expensive
    - No *in vivo* mode of operation



Time Series Data
Credits: ChaLearn Connectomics Challenge

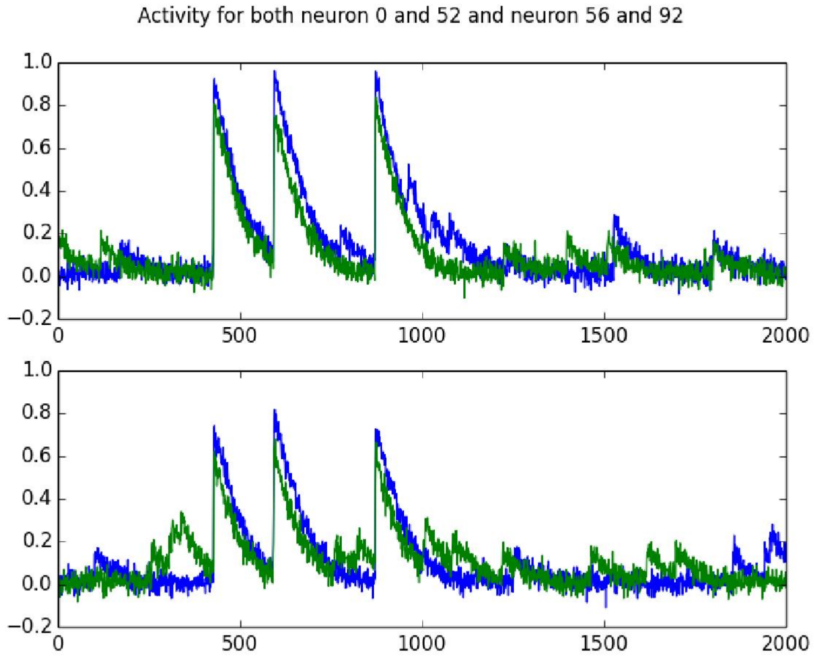
Causal Relations

The firing of a ***presynaptic*** neuron (**source** of directed links) is expected to affect the probability of firing of its ***postsynaptic*** neurons (**targets** of the directed links).

- Excitatory → the probability of firing will rise.

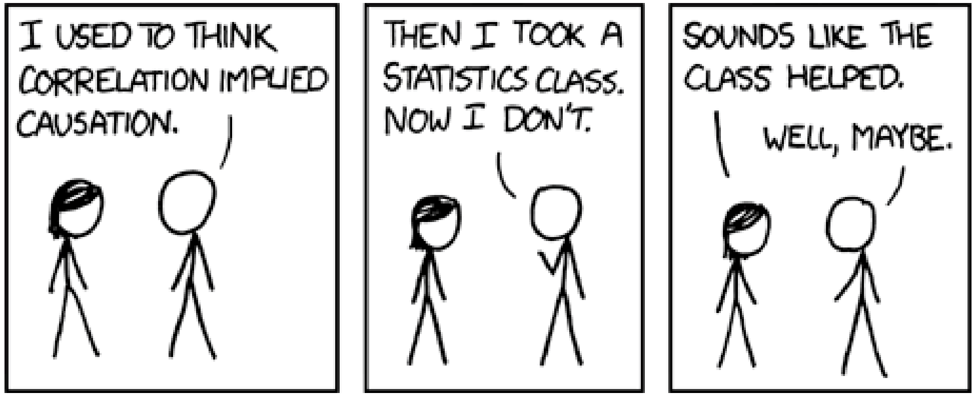
- Inhibitory → probability of firing would decrease.

We consider exclusive excitatory neuronal connectivity.

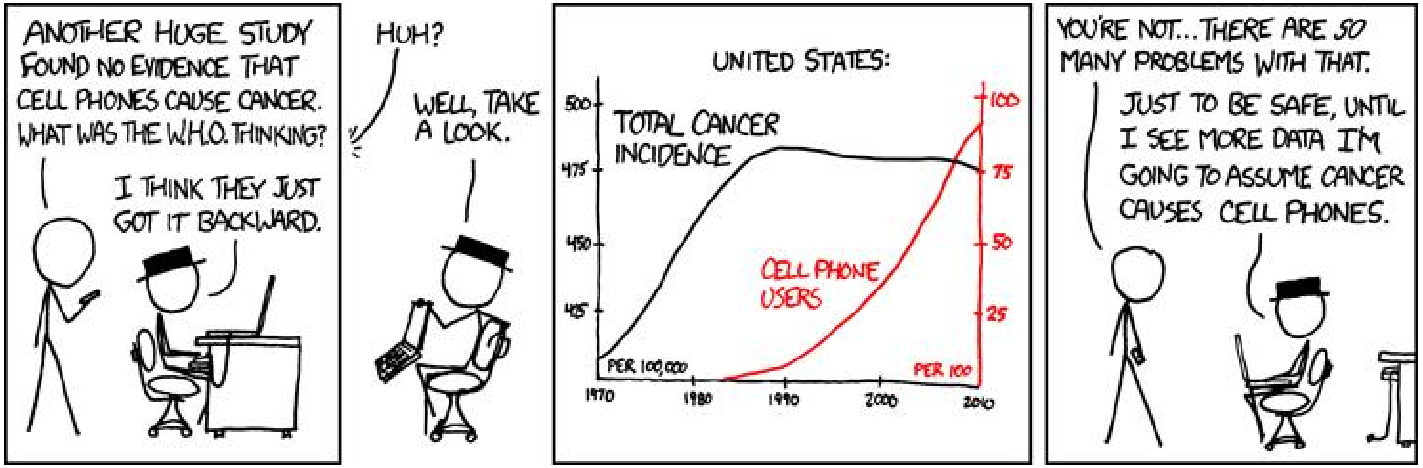


Time variation of neural activity.
Credits: Self

And we run into problems…

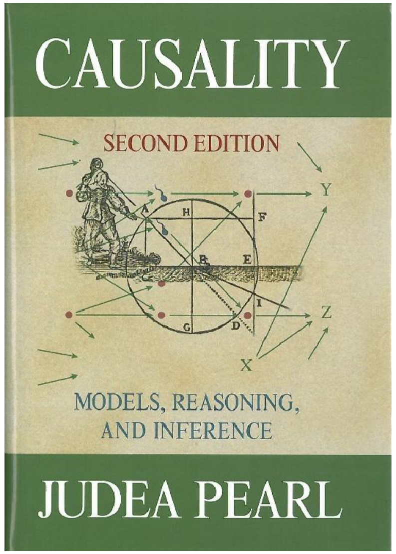


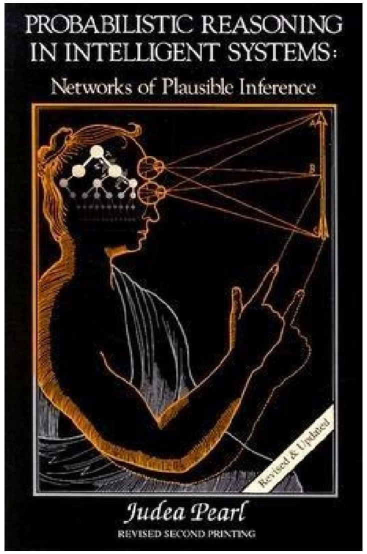
552: Correlation
Credits: **xkcd**



925: Cell Phones
Credits: **xkcd**

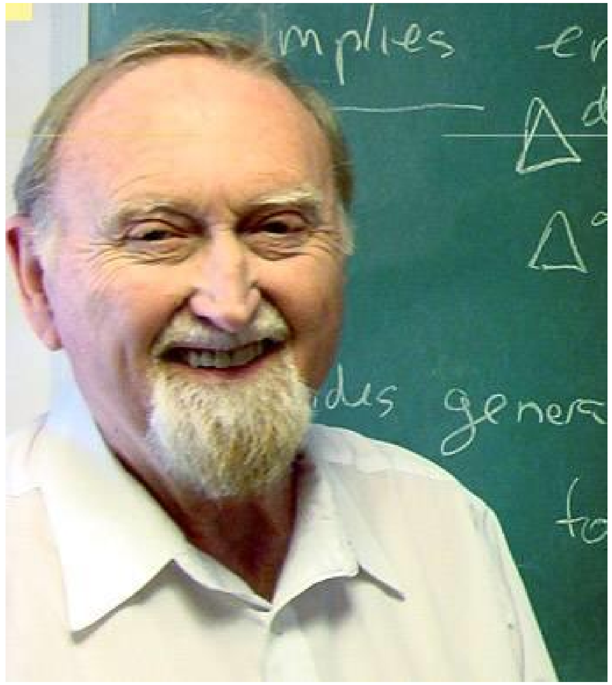
Fundamental Contributions





Credits: Amazon

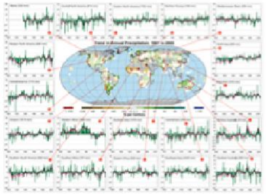
Judea Pearl
2011 **ACM Turing Award**
"for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning"

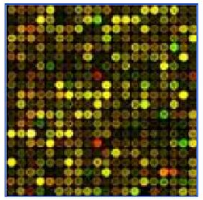


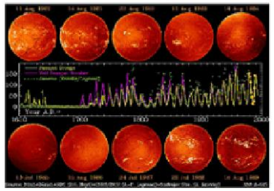
Clive Granger
2003 **Nobel Prize** in **Economics**
"for methods of analyzing economic time series with common trends"

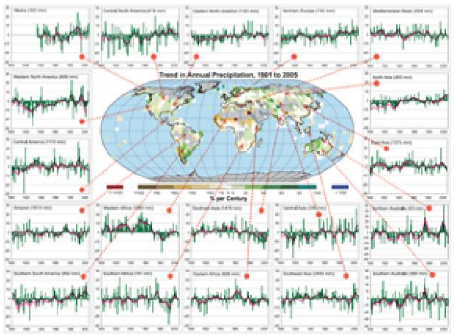
Time Series Data is Ubiquitous

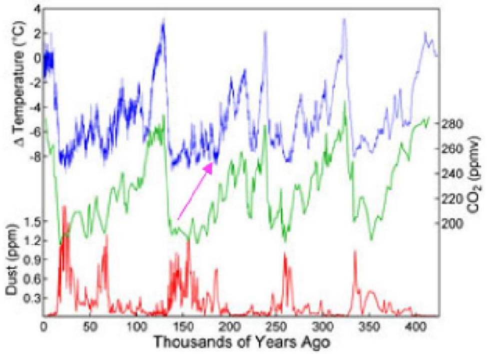
Discovery of Temporal Causal Relationships











Understanding Climate Change Data
Credits: **[CIKM Tutorial]**

Granger Causality

- A cause is prior to its effect.
- If $X \xrightarrow{Granger} Y$, then X^{past} should significantly help predicting Y^{future} via Y^{past} alone.
- Based on linear regression/prediction of time series (Granger 1969)

Two Principles

1. The cause happens prior to the effect.
2. The cause makes unique changes in the effect. In other words, the causal series contains unique information about the effect series that is not available otherwise.

Granger Causality

We perform two vector auto-regressions as follows:

$$Y(t) = \sum_{l=1}^{L} a_l Y(t-l) + \epsilon_1 \quad (1)$$

$$Y(t) = \sum_{l=1}^{L} a_l' Y(t-l) + \sum_{l=1}^{L} b_l' X(t-l) + \epsilon_2, \quad (2)$$

where L is the maximal time lag. We say X causes Y if eq (2) is statistically significantly better than eq (1).

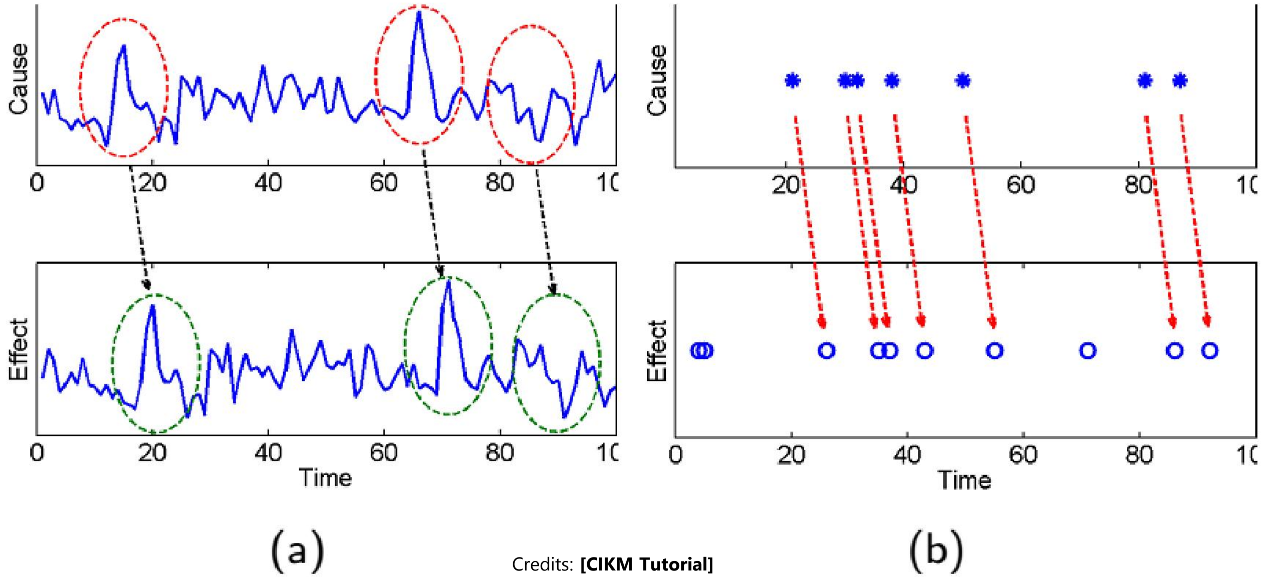
Notes on “Statistically Significant”

- Any statistic blind to hypothesis-space complexity will indicate a better fit when number of parameters are more.
- Need to **penalize** on basis of number of parameters used in the model.
- Seeming popular in literature, **Akaike information criterion**

$$AIC = 2k - 2\ln(L)$$

  - *k* is the number of parameters in the statistical model
  - *L* is the maximized value of the likelihood function for the estimated model.

Example



Credits: **[CIKM Tutorial]**

Figure : Illustration of the main principle behind Granger causality: in both examples the cause happens prior to its effect and its past values help predicting future values of the effect. (a) Plot of the values of two time series. (b) Plot of two point processes in which each event is shown with a mark at its happening time.

"Of course, many ridiculous papers appeared!" – Clive Granger, 2003 Nobel Lecture

Granger Causality - Chicken or Egg?

• This causality test is also can be used in explaining which comes first: chicken or egg. More specifically, the test can be used in testing whether the existence of egg causes the existence of chicken or vise versa.

• Thurman and Fisher (1988) did this study using yearly data of chicken and egg productions in the US from 1930 to1983.

• The results:

1. Egg causes the chicken.
2. There is no evidence that chicken causes egg.

Source: Lecture Slides from Econometrics II www.bauer.eh.edu

Granger Causality - Remarks

• Granger causality does not equal to what we usually mean by causality.

• Even if x_1 does not cause x_2, it may still help to predict x_2, and thus Granger-causes x_2 if changes in x_1 precedes that of x_2 for some reason (usually because of a third variable, missing in the model).

• <u>Example</u>: A dragonfly flies much lower before a rain storm, due to the lower air pressure. We know that dragonflies do not cause a rain storm, but it does help to predict a rain storm, thus Granger-causes a rain storm.

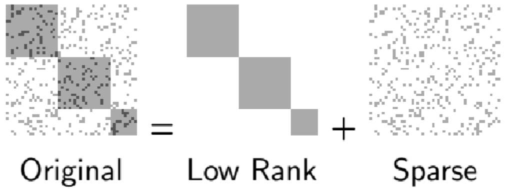
Granger Causality Extended

- **Vector-Auto-Regression (VAR)** based Granger causality test assumes linear dependence.
  - Non-parametric approach: Fourier or Wavelet transformation, Kernel methods
- How to handle **Latent Factors**?

Common Approach For Local Impact Latent Factor

1. Explicitly model the hidden variables in the graphical model.
2. Apply the Expectation Maximization algorithm to learn the structure.

Sparse plus Low-rank Decomposition For Global Impact Latent Factor



Credits: **[CIKM Tutorial]**

There are convex optimization algorithms to decompose a matrix into a sparse matrix and a low-rank matrix.

Overview of proposed methods

- Cross-correlation

$\gamma_{xy}(\tau) = \mathrm{E}[(x_t - \mu_x)(y_{t+\tau} - \mu_y)]$ where μ is the mean

The normalized version of cross correlation can be obtained by treating both x and y time series as vectors and dividing by the norms of x and y

To compute an index value for causality take the average γ over

$\tau = 0..m, m \in [0, T-1]$ and T is the total time steps for time series x and y

- Granger Causality

Fit:

$$y_t = a_t + \sum_{k=1}^{p} \alpha_k y_{t-k} + \sum_{k=1}^{q} \beta_k x_{t-k} + u_t \text{ and}$$

$$y_t = a_t + \sum_{k=1}^{p} \alpha_k y_{t-k} + e_t$$

Conduct an F-test to check the null hypothesis that $\forall k, \beta_k = 0$ by comparing the sum of squares of residuals RSS:

$F = \dfrac{(RSS_0 - RSS_1)/q}{RSS_1/(T-p-q-1)}$ where $RSS_1 = \sum_t \hat{u}_t^2$ and $RSS_0 = \sum_t \hat{e}_t^2$ The F statistic follows an F-distribution with DOF $(q, T\text{-}p\text{-}q\text{-}1)$

Overview of proposed methods

- Generalized Transfer Entropy

$$\mathrm{GTE}_{xy}(\tilde{g}) = \sum_{i,j,k} P(y^k_{t+1}, y^j_t, x^i_{t+1} \mid g_{t+1} < \tilde{g}) \log\left(\frac{P(y^k_{t+1} \mid y^j_t, x^i_{t+1}, g_{t+1} < \tilde{g}}{P(y^k_{t+1} \mid y^j_t, g_{t+1} < \tilde{g})}, \right)$$

where the subscript t denotes a time step and the superscripts i,j,k denote a possible value of the corresponding variable. g_t is the average fluorescence signal of the entire network and $\tilde{g}$ is a predefined threshold parameter

- Information Gain

Treat all time points of the time series as i.i.d
IG(y,x)=I(y)-I(y|x), where I can be any measure of impurity such as Gini index or Entropy
Gini index : $G(x) = 1 - \sum_i P(x^i)^2$

Transfer Entropy

Define the following information sets:

- $\mathcal{I}^\star(t)$ - the set of all information in the universe up to time t;
- $\mathcal{I}^\star_{-X}(t)$ - the set of all information in the universe excluding X up to time t.

Granger's definition of causality (1969, 1980). Given two time series X and Y, we say X causes Y if

$$\mathbb{P}[Y(t+1) \in A | \mathcal{I}^\star(t)] \neq \mathbb{P}[Y(t+1) \in A | \mathcal{I}^\star_{-X}(t)],$$

Definition (Schreiber, 2000)

$$T_{X \to Y} = \mathbb{H}\left(Y^t | Y^{\mathbf{t-1}}, X^{\mathbf{t-1}}\right) - \mathbb{H}\left(Y^t | Y^{\mathbf{t-1}}\right)$$

The amount of resolved uncertainty in future of Y by past values of X given past values of Y.

Properties

- Non-linear causation
- Non-stationary time series
- Conceptually more meaningful;
  since uses independence instead of correlation.

Limitations

- Only well-defined for two time series
- Computation of entropy from observations is challenging.

Relationship with Granger Causality (Barnett, 2009)

Transfer Entropy and Granger Causality are equivalent if the data is distributed according to multivariate Gaussian distribution.

Transfer Entropy

The average number of bits needed to encode independent draws of a r.v. **I** following a distribution **p(i)** is

$$H_I = -\sum_i p(i) \log_2 p(i)$$

Excess number of bits coded if a different distribution **q(i)** is used for the coding instead of the actual underlying distribution **p(i)**

$$K_{I|J} = \sum_{i,j} p(i,j) \log \frac{p(i\,|\,j)}{q(i\,|\,j)}$$

Excess amount of code produced by assuming that the two systems are independent

$$M_{IJ} = \sum p(i,j) \log \frac{p(i,j)}{p(i)p(j)}$$

Excess amount of code produced in coding "destination" variable **I** by assuming that the next state of the destination variable is independent of the "source" variable **J**

$$T_{J\to I} = \sum p(i_{n+1}, i_n^{(k)}, j_n^{(l)}) \log \frac{p(i_{n+1}\,|\,i_n^{(k)}, j_n^{(l)})}{p(i_{n+1}\,|\,i_n^{(k)})}$$

Dataset

- Data provided
  - 20 ms time series data for hundred/thousand neuron.
  - Coordinates of each neuron.
  - Ground truth (except for test set).
- The data has been provided under
  - Varying settings of sub-normal and super-active neural activity
  - Different values of clustering coefficient

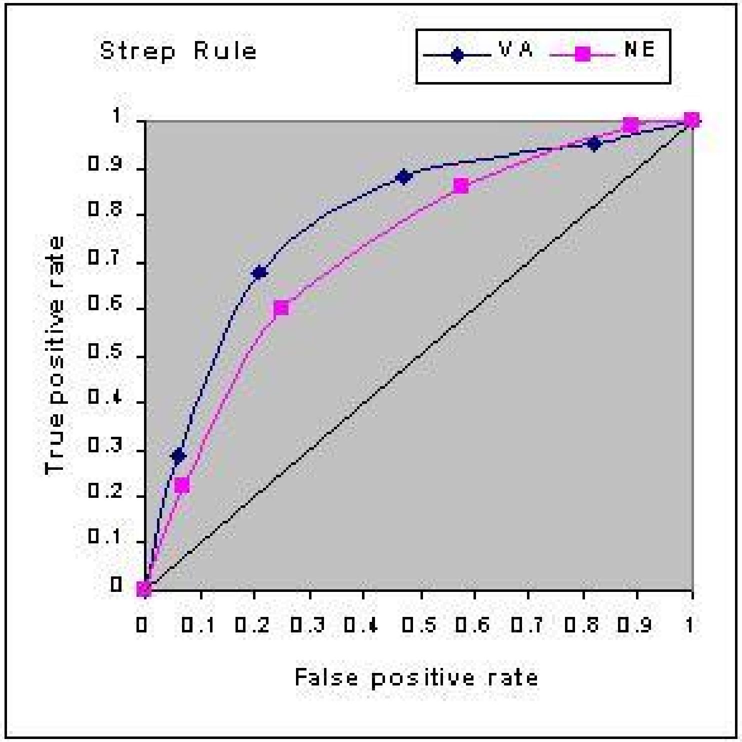
Evaluation Metric

- We return a numerical score between 0 and 1 indicating our "confidence" that there is a directed connection.

		Prediction	
		label 1	label 0
Truth	label 1	tp	fn
	label 0	fp	tn

- True positive ratio = tp/(tp+fn) = Recall(Label 1)
  False positive ratio = fp/(fp+tn) = 1-Recall(Label 0)

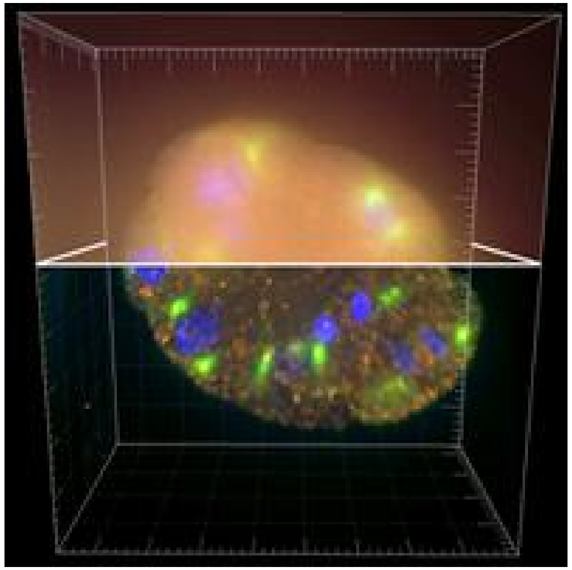
- <u>Aim:</u> Maximize area under the ROC curve.



ROC curve
Credits: Wikipedia

Challenges

- Possibly high **signal-to-noise** ratio.
    - Solution: Use signal deconvolution methods.
- Robustness with respect to **Collective Synchrony.**
    - Solution: Conditioning on global mean activity.
- Insufficient **resolution** of time series data.



Credits: 3D Deconvolution Microscopy Challenge

References

[1] **(Challenge)** ChaLearn Connectomics Challenge → http://connectomics.chalearn.org/

[2] **(CIKM Tutorial)** Causality Analysis in Large-scale Time Series Data, Yan Liu, CIKM 2013 Tutorial

[3] **(Stetter, 2012)** Model-Free Reconstruction of Excitatory Neuronal Connectivity from Calcium Imaging Signals, Olav Stetter, Demian Battaglia, Jordi Soriano, Theo Geisel, PLOS Computational Biology, August 2012, Volume 8.

[4] **(Nolte-Muller, 2010)** Localizing and Estimating Causal Relations of Interacting Brain Rhythms. Guido Nolte, and Klaus-Robert Müller, Front Hum Neurosci. 2010; 4: 209, 2010.

[5] **(Orlandi, 2013)** Transfer Entropy reconstruction and labelling of neuronal connections from simulated calcium imaging, Orlandi, J.G., Stetter, O., Soriano, J., Geisel, T., and Battaglia, D. (2013). ArXiv 1309.4287.

QA &

Suggestions

Thank You!