Connectomics!

Neural Structure Reconstruction from Fluorescence Imaging

Course Project – CS 365A

Karan Singh, Pankaj Gupta
Indian Institute of Technology, Kanpur

April 24, 2014

Abstract

Understanding the topology underlying neural connections and the associated changes induced by defects is often deemed a promising approach to deciphering the puzzle of neuro-pathologies, eg. epilepsy, Alzheimer's disease.

In this work, we would like to reconstruct the network structure of the brain from time-series activity of the neurons monitored through calcium-based fluorescence imaging. Since the connections are directed in nature, the fundamental obstacle the task entails is that of inferring statistical causality from time-series data. We investigate various algorithmic techniques used for inferring statistical causality. In the process, we come across some interesting features exhibited by biological neural networks. We also successfully mitigate a few challenges particular to the problem.

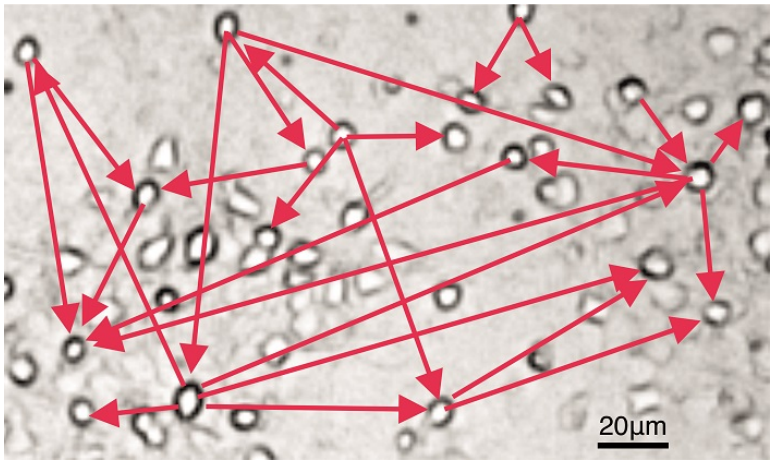

Figure 1: Network discovery of live neurons through fluorescent calcium imaging. Neurons the pictures are white dots. The connections recovered are represented as red arrows. Credits: [4].

Preface

It is good time to be doing Computational Biology[1]. Imaging (and in general, signal acquisition) techniques have advanced to a point where the rate of generation of data now exceeds our ability to process it. Automating the process of drawing useful, robust inferences from such high-volume data is an essential and challenging task.

Recognizing this challenge, there has been an increased focus on use of statistics and machine learning in these domains. For example, a lot of MOOC platforms today offer courses such as Statistical Analysis of fMRI Data[2] and Data Analysis for Genomics[3]. We tackle a very relevant problem of reconstruction of neural structure from fluorescence imaging of neural activity.

Acknowledgements

Every year about half-a-hundred unsuspecting students take up the course on Artificial Intelligence at the Indian Institute of Technology, Kanpur. For most, the course offers a very different experience – being forced head-on against a research problem – in a sharp contrast to the other courses, which offer a mild and guided introduction to the subject. The experience is akin to what happens when a novice tries riding a bicycle for the first time. There are scratches here and there; but, the journey is exciting.

The authors wish to thank Prof. Amitabha Mukerjee and the course staff – M Seetha Ramaiah and Sunakshi Gupta – for providing the students (the authors included) with this enthralling "adventure" and an unadulterated introduction to hands-on research.

Karan Singh
karansng@iitk.ac.in
Pankaj Gupta
pankajg@iitk.ac.in

[1] We do admit guilt on using the term incorrectly. For a finer distinction between Computational Biology, Bioinformatics and Biostatistics, see [1].

Contents

1

Introduction

"You, your joys and your sorrows, your memories and your ambitions, your sense of identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules. as Lewis Carroll's Alice might have phrased it: You're nothing but a pack of neurons."

– Francis Crick, *The Astonishing Hypothesis(1994)*

1.1 Objectives

The aim of the project is to predict the excitatory connection between neurons given the time series data of neuronal activity captured through fluorescence imaging. Understanding the neural connection-topology can be very helpful in understanding the structure of brain and the unravelling the mysteries of brain's learning capacity. It would also facilitate the study of causes and cures of neuro diseases such as epilepsy, Alzheimer's disease.

1.2 Some Neurobiology

Neurons are cells which constitute the nervous system. They are electrically excitable and can transmit electrical and chemical signals. The transfer of signal occurs through synapses. Dendrites receive the signal through synapse. Axons carries the signal through a distance and then transmit it. When the action potential reaches the axon ending, neurotransmitters are released into the synaptic gap and facilitate the transfer of information.

On an average, the human brain contains about 100 billion neurons. Coupled with fact that the average number of synaptic connections are about seven thousand per neuron, it implies that the human neural network is gigantic compared to the typical number of neurons used in artificial neural networks.

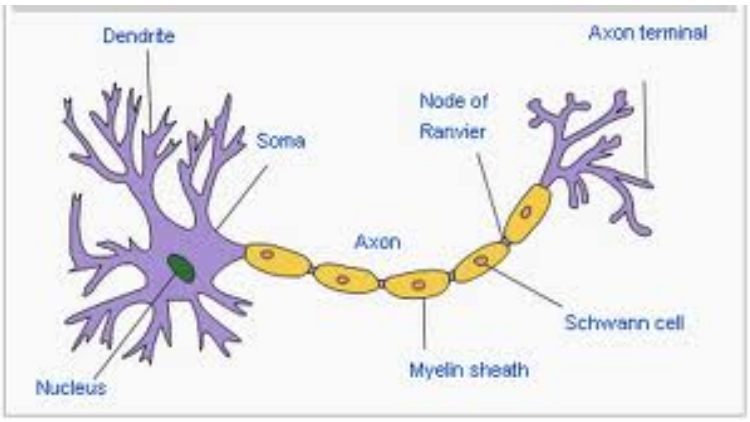

Figure 1.1: Structure of a typical neuron. Credits: Wikipedia

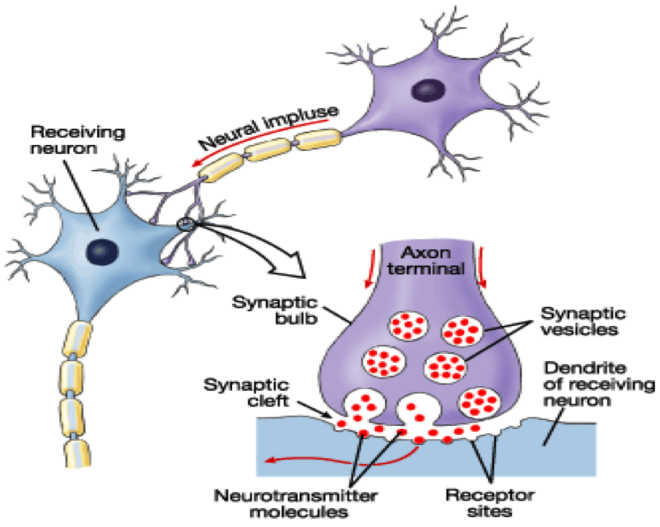

Figure 1.2: Transmission of signal between two neurons via synapse. Credits: Psychology1003 `www.studyblue.com`

1.3 Fluorescence Imaging

Fluorescence imaging is a technique which takes advantage of fluorescence nature of calcium indicators to detect the status of calcium in brain and other neuron containing tissues. The transmission of signal from one neuron to the other induces calcium influx in the synapse. This potential influx can be monitored using calcium indicator dyes. The technique can be applied to monitor calcium activities, both *in-vivo* and *in-vitro*. However, it has a drawback of low signal to noise ratio. Hence elementary signals are hard to detect.

1.4 Existing approaches and Problems

Some contemporary techniques employed for predicting the neural connectivity are the techniques of Axonal Tracing and Electron Microscopy. Tracing at cellular level is a very difficult task and very much due to the large number of neurons and synaptic connections, Axonal Tracing do not scale enough to be of practical value. On the other hand, Electron Microscopy is very costly and cannot be used for any type of in-vivo operation.

1.5 Task Description

The data used in the present work comes from a Kaggle competition[4]. A description of the same follows – the data consists of:

- Time series of neural activities obtained from fluorescence signals sampled at 20ms intervals.
- (X, Y) coordinates of neurons. Assuming that each neuron spans a 1 mm^2 area.

The output format consists of the confidence levels associated with the presence of directed links between pairs of neurons, so as to be meaningful when evaluated against an AUC score.

The data provided to us is a synthetic data[4] generated using a realistic model of neuron and fluorescent imaging. The model has also simulated limitations and defects of the calcium imaging technology i.e limited time resolution and light scattering artifacts (which ,means that activity of given neuron influences the measurements of nearby neurons).

2

Statistical Causality

> *At that time, I had little idea that so many people had very fixed ideas about causation, but they did agree that my definition was not "true causation" in their eyes, it was only "Granger causation." I would ask for a definition of true causation, but no one would reply. However, my definition was pragmatic and any applied researcher with two or more time series could apply it, so I got plenty of citations. Of course, many ridiculous papers appeared.*
>
> – Clive Granger, *Nobel Lecture*

The task of reverse engineering neural structure topology from the activity data seems a very promising venture to by-pass the practical infeasibilities that bind the current methods. The goal of this challenge is to infer directed connections between neurons from time-series patterns of neural activity. Such a directed graph may be interpreted as a causal network. Further, neurons have complex temporal patterns of activity. The challenge can therefore be viewed as that of causal structure reconstruction from time series data[4]. Moreover, such a view seems to be well supported by the mechanism[11] of neuron firing, especially since we restrict our work to excitatory neural connections. It has been empirically observed that such connections tend to enhance the probability of firing of the 'effect' neuron if the 'cause' neuron has fired in the recent past. This draws close parallels to the events like – analyzing if CO_2 causes global warming – which are classical examples involving statistical causality.

2.1 Why is it difficult?

Statistical causality is often seen as a tricky subject to deal with in the Statistics community. The challenge becomes even more difficult in case of

observational studies[1]. One has to deal with effects such as[13]

- Reverse causation – Anticipating the effect, the cause-inducing agent changes her actions.

Figure 2.1: SMBC (Saturday Morning Breakfast Cereal) Comics – Reverse Causality Example

[1]For an explanation on why this is indeed the case, see [12] which cleanly draws a contrast between randomized control experiments and observational studies.

- Regression to the mean – If the first measurement of a random variable results in an extreme value, the second is likely to be closer to the true mean than the estimated mean value.
- Common cause based confounding – A common cause results in 2 effects, which may be incorrectly be interpreted as cause-effect pairs.
- Differential selection based confounding – In an observational study, the composition of groups tends to be such so as to favor a particular interpretation.

The xkcd comics below do an excellent job at explaining the challenges.

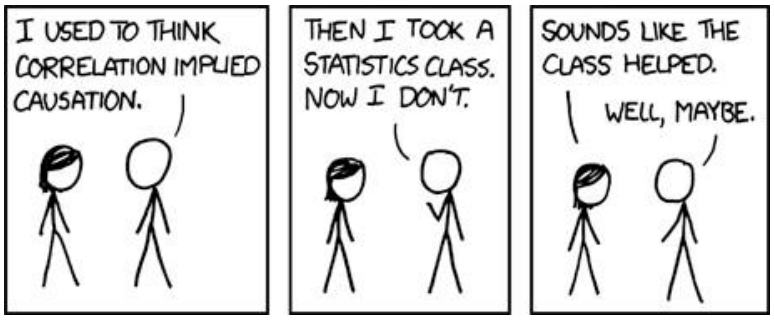

Figure 2.2: xkcd 552: Correlation

In the above comic strip, Cueball is trying to explain Megan that he earlier thought correlation implies causation but after he took the statistics class he doesn't thinks so. When Megan inquired that statistics class was the causation of his change of belief, he himself is confused! Taking statistics class was perhaps correlated to his change of belief, but fails to serve as an evidence for it.

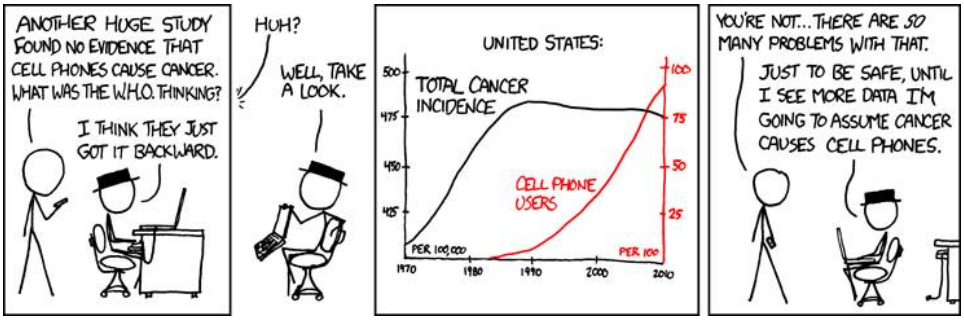

Figure 2.3: xkcd 925: Cell Phones

In the above comic strip, Cueball tells Black Hat about another study which showed that cell phones causes cancer. Black Hat tells him that its the other way. He had a plot which clearly showed that after the cancer rose, the cell phones rose. Hence, Black Hat believes that cancer is the cause cell phones. Black Hat's belief is based on the principle that if X occurs before Y, then X is the cause of Y which is not always true. This principle forms the basis of Granger causality (which we explain in the section that follows).

There has been much philosophical debate on whether one can make claims about causality without disturbing the system itself, that is without intervening in the working of the system. But, in the current work, we set these topics aside and look at what we can do given the data at hand – we look at functional or predictive causality.

2.2 Major Lines of Work

The field of statistical causality has seen two major lines of work – while one borrows from econometrics and the other from computer science. We choose to briefly discuss them below.

2.2.1 Conditional Independence

Judea Pearl came up with the idea of representing interacting variables as vertices on a directed acyclic graph, with the interpretation that edges go from 'cause' to 'effect'. Now, in such a graph, if two variables do not have an edge between them, it can be shown that they are conditionally independent given that the other vertices hold fixed values. Such vertices, which are conditionally independent, do not have any cause-effect relationship.

2.2.2 The works of Clive Granger

Granger causality was one of the first ideas introduced for inferring causality in time series data. The history goes back to Nobel laureate economist Clive Granger who started looking at how causal relations could be discovered in time-series data in econometrics[5]. The idea is to see how better can one do by predicting the current value of the 'effect' based on the past values of the 'cause', as compared to without the past values of 'cause'.

The careful reader will note that the definition is not free of problems. The author (unashamedly!) generously borrow the following example from [7] – "A dragonfly flies much lower before a rain storm, due to the lower air pressure. But, that does not mean that through the activity of flying low

the dragonfly causes the rain storm."

While the assertion is true, Granger causality was aimed at only being a functional measure of causality. So if the past of $\mathbf{X}$ helps in predicting the values of $\mathbf{Y}$, $\mathbf{X}$ does functionally cause $\mathbf{Y}$. The quote at the beginning of the chapter by Granger does better defending the idea.

3

Exploratory Data Analysis: Insights and Discoveries

In this chapter, we develop some insights into the data. Apart from forming the core of our pre-processing steps and boosting the scores, they also reveal interesting features of biological neural networks. A few insights help us pin-pointing the central difficulties associated with the task. We successfully mitigate a few of these challenges.

3.1 Neural Activity

The first observation is that every spike begins a rapid increase in observed activity, followed by a slow decay. However, for the purpose of inferring causality, it suffices to look at spiking times instead of neuron behaviour. The prolonged period of decay may also adversely affect the performance of the reconstruction algorithm. We see the above stated behavior in Figure 3.1. Hence, to get an estimate of spiking times. We first differentiate the

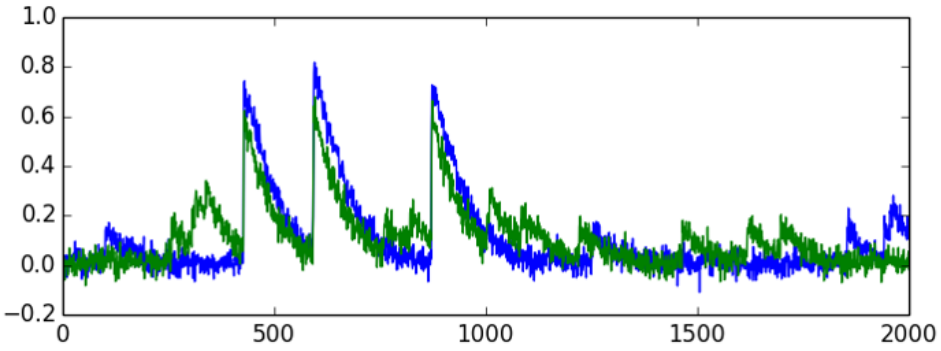

Figure 3.1: Neural Activity vs. Time

time-series data, and then mark each point as a spike where the value of the first derivative exceeds a threshold.

$$\Delta(X_n) = X_{n+1} - X_n \tag{3.1}$$

3.2 Collective Synchrony

This section talks about a very interesting discovery about biological neural networks. We consider the following set of observations. Given a table of time-series data of 100 neurons, project it onto the time axis, thereby, computing the net amount of neural activity at given time point. The plot below demonstrates this. One could infer that at some time points, either

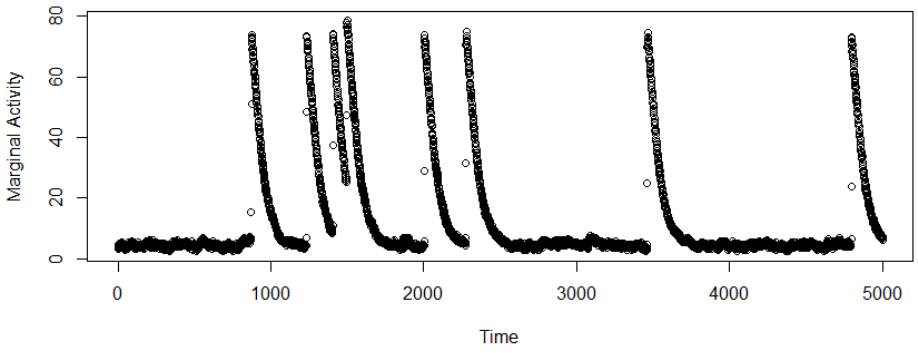

Figure 3.2: Spikes in neural activity indicate collective synchrony.

all or a handful of neurons get hyper-active which results in a very high net amount of neural activity. It is also useful to note that the onset is sudden and happens for all involved at about the same time point. To confirm our intuition, we plot the heatmap of neural activity. Now, we are in a position to make a stronger statements. There are times at which **all** the neurons get activates resulting in a very high net activity, which happens with a sudden and coherent offset. This phenomena is commonly called "network burst"[17] or "collective synchrony"[16] in the literature.

In such regions of high activity, inferring the directed connections between neurons becomes a challenging task, because there are multiple 'cause' neurons firing at the same time for each 'effect' neuron. Hence, the connection inferred during this period can be spurious. To get past this problem, we ignore the data points from these high activity regions in the data-set. This change, coupled with the same-bin interaction, boosts the performance of the algorithms in the next section by about **39%**.

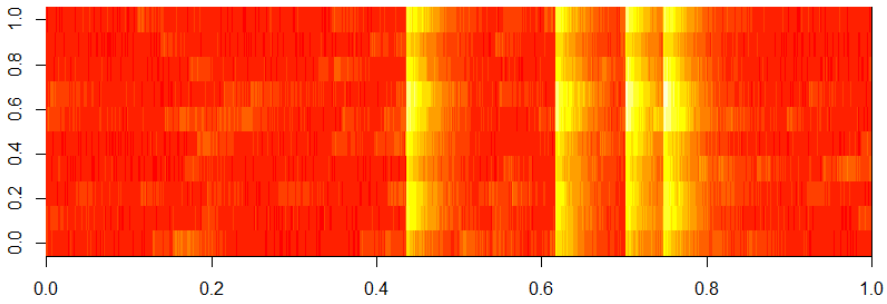

Figure 3.3: Heatmap of Neural Activity vs. Time

3.3 Same-bin Interaction

We also note that the time series resolution is insufficient – that is both the 'cause' and 'effect' can fire within the sampled time point. This can be demonstrated by some very simple back-of-the-envelope calculation. The time series data has been sampled at intervals of 20 ms. The speed at which impulses in neurons travel is about 100 ms^{-1}[14]. This means for the time series resolution to be sufficient the neurons, on average, must be 2m apart which is clearly not the case. The authors conjecture that a resolution of about 0.1 ms would perhaps serve the purpose.

To counter the problem, we introduce a change in all of our algorithms taking into account the present value of 'cause' in addition to the other factors when predicting the present value of the 'effect'. This takes into account the interactions taking place between cause effect-pairs in the same time-bin. As a result of our pre-processing steps, the AUC score of the cross-correlation measure goes up from **0.58** to **0.81**.

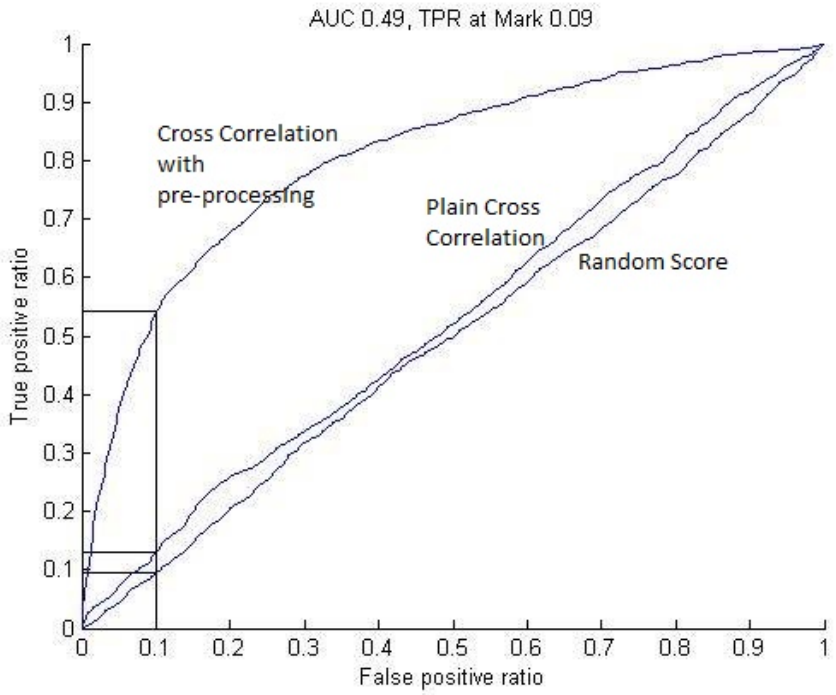

Figure 3.4: Pre-processing increases AUC score for Cross Correlation

4

Algorithms for Causality Inference

In this chapter, we briefly review various algorithmic techniques used for establishing causal relations in our work. The emphasis is on the assumptions each makes, hence, contrasting the applicability of one against another.

4.1 Cross-correlation

We begin with the simplest measure – cross correlation. The standard correlation between 2 random variables measures how frequently do both variables attain similar values. The cross correlation extends the idea; and asks how frequently does the 'effect' get activated given that the 'cause' was active in the near past.

$$XC_{Y \to X} = \max_{\Delta t = 0 \ldots t_{max}} corr(X_S, Y_{S-\Delta t}) \tag{4.1}$$

In spite of its inherent simplicity, cross correlation happens to be the de facto standard in practice today. As we shall see in the next chapter, our study, to a good extent, justifies this with cross correlation achieving an AUC score of 0.81 outdoing many sophisticated information-theoretic measures.

4.2 Granger Causality

As described in the previous chapter, the Granger causality involves doing a linear fit. Mathematically, the procedures involves running a multivariate linear regression on the current values of 'effect' with the values of 'past' values of 'effect' as covariates in the first go. In the second round, we do a multivariate linear regression on the current values of 'effect' with the values of 'past' values of 'effect' and the past values of the 'cause' as covariates. If the second regression results in a significantly better fit, one concludes

that the past values of 'cause' do help in predicting the current values of the 'effect' – thereby, enforcing that the 'cause' does indeed cause the 'effect'.

$$x_t = \sum_{l=1}^{k} a_l^0 x_{t-l} + \eta_t^0 \tag{4.2}$$

$$x_t = \sum_{l=1}^{k} a_l^1 x_{t-l} + \sum_{m=1}^{k} b_m^1 y_{t-m} + \eta_t^1 \tag{4.3}$$

$$GC_{Y\to X} = \log \frac{\Gamma_{0,0}^0 + \Gamma_{1,1}^0}{\Gamma_{0,0}^1 + \Gamma_{1,1}^1} \tag{4.4}$$

We use standard notations – Γ_i is the covariance matrix that results from each of the fits. There are a handful of assumptions involved here. Leaving aside any philosophical debate, the idea of taking a linear fit is a pretty strong one and limits the applicability of the measure. Another serious issue with this approach is that the method fails in presence of unaccounted confounding factors. Nevertheless, the idea lends itself to simple and useful modifications which deal with the stated problems.

4.3 Mutual Information

Mutual information is a measure of how information two random variables share – the degree of their mutual dependence. It is defined as the difference in the entropy values assuming the variables are independent and another case taking into account their joint distribution. It is easy to see that if the variables are mutually independent, the value is zero. To take into account the temporal lag, we measure the mutual information of the present of 'effect' with the past of 'cause'.

$$MI_{Y\to X} = \max_{\Delta t=0\ldots t_{max}} \sum_n P(X_n, Y_{n-\Delta t}) \log \frac{P(X_n, Y_{n-\Delta t})}{P(X_n)P(Y_{n-\Delta t})} \tag{4.5}$$

The primary reason why mutual information is often deemed unfit for inferring causal relation is its symmetry (sans the temporal lag) among the two variables. In particular, not considering the temporal lag introduced, $MI(X,Y) = MI(Y,X)$.

4.4 Information Gain

Information gain is a general framework for looking at the causal measures. Given any function $I(X)$ which is a representative of some measure of information content of a variable, we compute the difference between the marginal and conditional distributions.

$$IG_{Y\to X} = I(X) - I(X|Y) \tag{4.6}$$

4.4.1 Entropy based Gain Measure

An obvious choice of $I(X)$ is the entropy, which yields many mathematically aesthetic properties. The entropy content of a random variable can be viewed as the expected number of bits required to report an event that the variable can attain. In some sense, it is a measure of compressibility. It is instructive to note that the entropy is maximum when the distribution is uniform – since you can not dis-count occurrence of some event. On the other hand, for a skewed event, the entropy can be quite low.

$$I(X) = \sum_{i=1}^{n} P(X_i) \log \frac{1}{P(X_i)} \tag{4.7}$$

4.4.2 Gini Index based Gain Measure

Another popular choice of $I(X)$ is the Gini Index, which is computed as the formula below. The Gini index is a measure of how the data is spread – dispersion. Our experimental analysis indicates that the Gini index and entropy measures have comparable performance.

$$G(X) = 1 - \sum_{i=1}^{n} P(X_i)^2 \tag{4.8}$$

4.5 Generalized Transfer Entropy

Our weapon of choice in the present work is Transfer Entropy introduced by Thomas Schreiber[8]. Since it involves conditional probability distribution, it is clearly asymmetric, in contrast with mutual information. Hence, it is capable of representing directional information. All the other methods surveyed up till this point assume the joint probability distribution does not change with time. It can be shown that the Transfer Entropy does not make any such assumption[8], while retaining the capability of representing non-linear measurements. Being devoid of such an assumption allows Transfer Entropy to ignore the common history and up to a certain extent, even, the confounding factors.

$$TE_{Y \to X} = \sum_{n} (P(X_{n+1}, X_n^{(k)}, Y_n^{(k)}) \log \frac{P(X_{n+1} | X_n^{(k)}, Y_n^{(k)})}{P(X_{n+1} | X_n^{(k)})}) \tag{4.9}$$

4.6 Combined Linear Model

At this point, for each pair of neurons, we have the scores due to each of the measures describe above. We ask the question if it is possible to combine the features to be able to better predict the causal relations. To do this,

we fit a linear model, treating the scores obtained as features, on some dataset for which the ground truth is given. This allows us to predict causal relations for test data. Not very surprisingly, the linear fit learnt has a large coefficient for Transfer Entropy measure, which it is the most successful of all measures. However, as we shall see in the next chapter, this does not result in a significant improvement compared to Transfer Entropy.

5

Results

The results obtained using various algorithms are summarized in the table below, along with the ROC curves. The ROC – Receiver Operator Characteristic – curve yields a value of 0.5 on the random score method, as is theoretically expected. A few remarks are due.

Algorithm	AUC
Generalized Transfer Entropy	0.83
Combined Linear Model	0.83
Cross Correlation	0.81
Information Gain (Gini)	0.78
Information Gain (Entropy)	0.76
Mutual Information	0.75
Granger Causality	0.49
Random Score	0.50

Table 5.1: Final results.

- Generalized Transfer Entropy seems to do the best. This can be attributed to it making fewer assumptions on the data – for example, not assuming that the joint probability distribution is time invariant.
- The combined linear model does as good as the Generalized Transfer Entropy. Not surprisingly, it also learns a large coefficient for GTE scores, when compared against other features.
- Granger Causality proves to be just as bad as making random guesses. The assumptions that the cause effect relationship can be explained by a linear model does not hold good in the data. We think that this is primary reason why Granger Causality fails so badly.

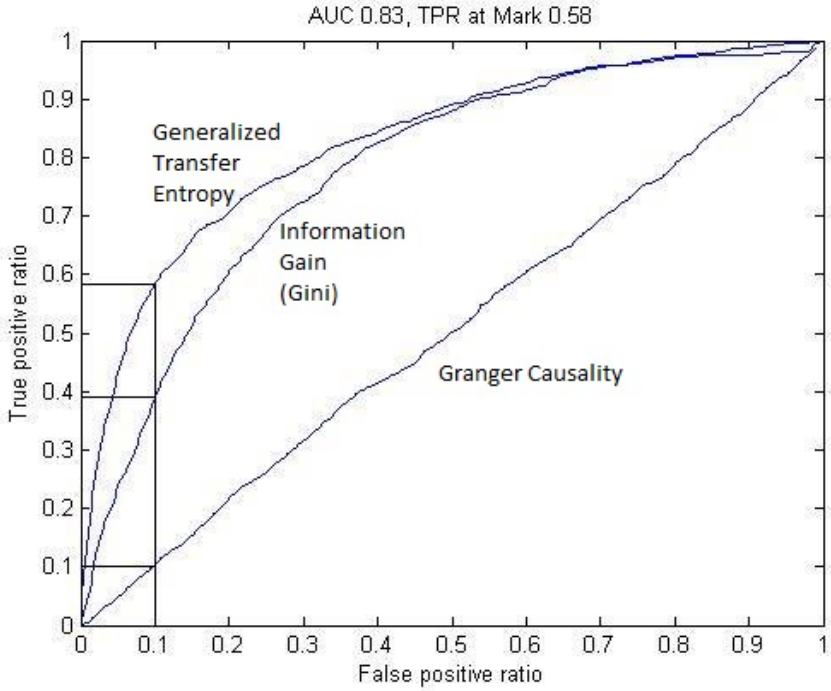

Figure 5.1: ROC Curve for {GTE, Information Gain (Gini), Granger Causality}

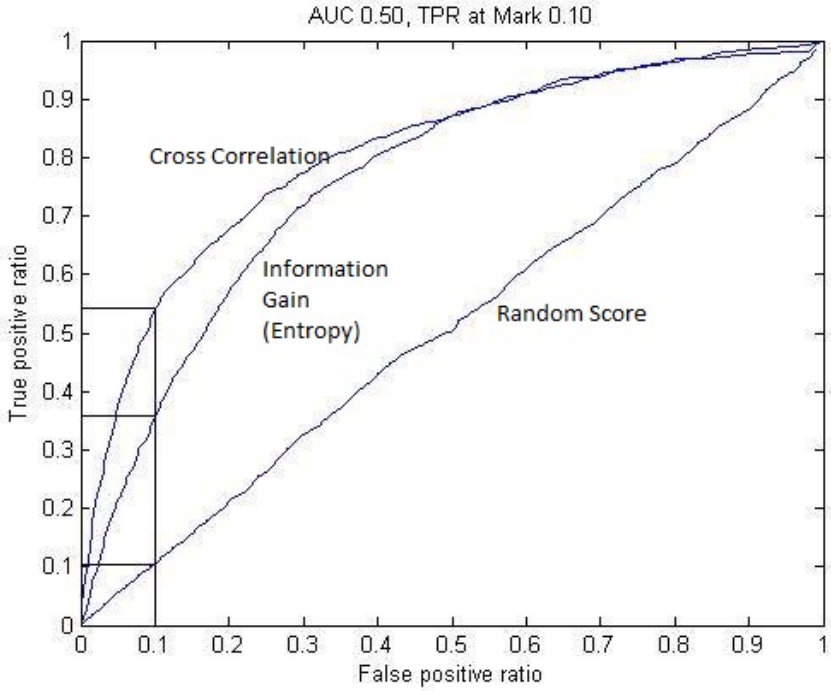

Figure 5.2: ROC Curve for {Cross Correlation, Information Gain (Entropy), Random Score}

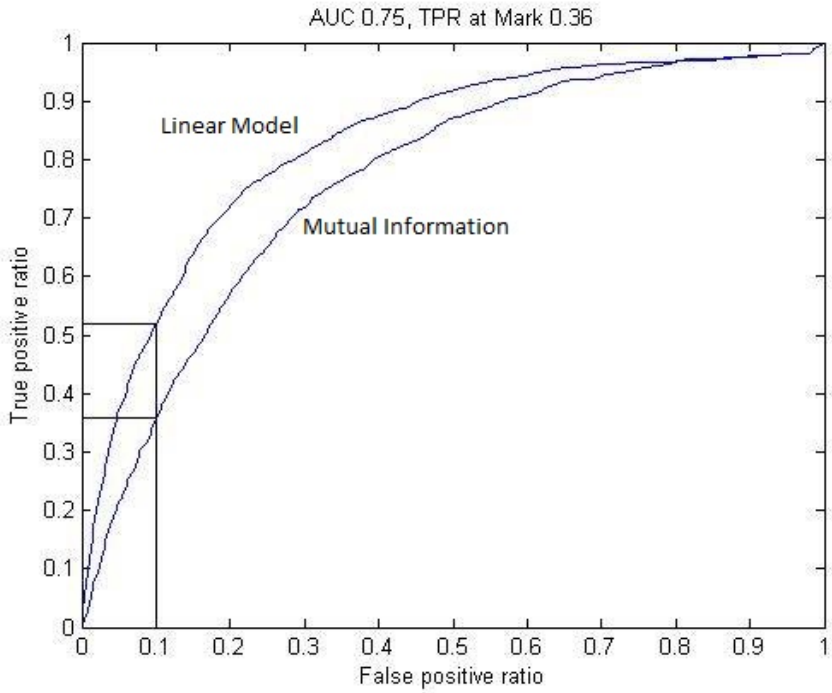

Figure 5.3: ROC Curve for {Linear Model, Mutual Information}

6

Future Work

The authors would like to state that the work done is no way complete and there are many (possibly easy to do) small tweaks as well as alternative approaches possible. A few promising ones are summarized below.

- **Complex Models:** With all the features (scores) computed, one could look at more complex models than linear fits, like SVMs and neural nets. While one has to be careful enough not to overfit, the chances of overfit are minimal – since the data is low-dimensional ($6 - 10$ features) and the number of samples are many (about 10^4).
- **Blind Deconvolution Techniques:** The data suffers from a low signal-to-noise ratio. In such scenarios, to counteract light scattering effects, one could utilize one of the available blind deconvolution techniques like [9].
- **Sparsity Inducing Models:** All of our approaches take into local factors – that is for deciding if there is a directed connection from **A** to **B**, we only take into account the time-series data of **A** and **B**. This may possibly be a major limitation. One way to mitigate this it to look at sparsity inducing reconstruction algorithms. These have been studied to some extent in the context gene regulatory networks[10]. Another idea is to resort to bayesian machine learning with sparsity inducing priors.
- **Speeding up existing algorithms:** Improving upon run time of existing algorithms will a great service to the community allowing for researchers to look at more expensive algorithms. As of now, the size of a data set serves a major obstacle to using more sophisticated algorithms.

Bibliography

[1] *NIH working definition of bioinformatics and computational biology.* Biomedical Information Science and Technology Initiative. 17 July 2000.

[2] *Coursera – Statistical Analysis of fMRI Data* John Hopkins University `https://class.coursera.org/fmri-001`

[3] *edX – Data Analysis for Genomics* Harvard `https://www.edx.org/course/harvardx/harvardx-ph525x-data-analysis-genomics-1401`

[4] *ChaLearn Connectomics Challenge: From Imaging to Connectivity* `http://connectomics.chalearn.org/`

[5] *Granger, C. W. J.* Investigating Causal Relations by Econometric Models and Cross-spectral Method *Econometrica 1969*

[6] *Yan Liu* Causality Analysis in Large-scale Time Series Data. *CIKM 2013 Tutorial* `http://www-bcf.usc.edu/~liu32/cikm2013.htm`

[7] *Rauli Susmel* Econometrics II – Lecture Slides `http://www.bauer.uh.edu/rsusmel/phd/ec2-6.pdf`

[8] *Thomas Schreiber* Measuring Information Transfer. *Physical Review Letters (2000)*

[9] *Joshua Vogelstein* Optimal OPtical Spike Inference `https://github.com/jovo/oopsi`

[10] *Amin Emad, Olgica Milenkovic* CaSPIAN: A Causal Compressive Sensing Algorithm for Discovering Directed Interactions in Gene Networks *PLOS ONE 2014*

[11] Excitatory Synapse *Wikipedia* `http://en.wikipedia.org/wiki/Excitatory_synapse`

[12] Observational Study *Wikipedia* `http://en.wikipedia.org/wiki/Observational_study`

[13] *Phil Dawid* Statistical Causality – Talk `http://www.statslab.cam.ac.uk/~apd/`

[14] *David Kraus* Concepts in Modern Biology.

[15] How fast does the human brain compute, and how much information can it store? `http://www.ualberta.ca/~chrisw/howfast.html`

[16] *Olav Stetter, Demian Battaglia, Jordi Soriano, Theo Geisel* Model-Free Reconstruction of Excitatory Neuronal Connectivity from Calcium Imaging Signals

[17] *J-P Eckmann, Shimshon Jacobi, Shimon Marom, Elisha Moses, Cyrille Zbinden* Leader neurons in population bursts of 2D living neural networks