

FAST ALGORITHM FOR NEURAL NETWORK RECONSTRUCTION

Sean Bittner^{1,2}, Siheng Chen^{1,2}, Jelena Kovačević^{1,3,2}

¹Dept. of ECE, ²Center for Bioimage Informatics, ³Dept. of BME,
Carnegie Mellon University,
Pittsburgh, PA, USA

ABSTRACT

We propose an efficient and accurate way of predicting the connectivity of neural networks in the brain represented by simulated calcium fluorescence data. Classical methods to neural network reconstruction compute a connectivity matrix whose entries are pairwise likelihoods of directed excitatory connections based on time-series signals of each pair of neurons. Our method uses only a fraction of this computation to achieve equal or better performance. The proposed method is based on matrix completion and a local thresholding technique. By computing a subset of the total entries in the connectivity matrix, we use matrix completion to determine the rest of the connection likelihoods, and apply a local threshold to identify which directed connections exist in the underlying network. We validate the proposed method on a simulated calcium fluorescence dataset. The proposed method outperforms the classical one with 20% of the computation.

Index Terms— connectivity analysis, nerves, machine learning

1. INTRODUCTION

Discovering the excitatory synaptic connections between individual neurons gives insight into computation at the lowest level of brain function. Understanding the topology of the human brain at the resolution of a single neuron is important for an understanding of structure and function of the brain [1].

To learn more about the brain, the connectivity of neural cultures¹ is studied. Determining this connectivity by axonal tracing, which consists of tracing the axon of each neuron to determine each point of synaptic connection, is challenging because of the huge number of neurons. For example, the human neocortex has roughly 20 billion neurons with 7,000 synaptic connections per neuron [2]. To study the human brain, even in smaller segments, an alternative to the labor intensive process of axonal tracing is necessary.

A classical approach to predicting the connectivity of neural networks is to collect and analyze their time-series signals

of the electro-chemical activity. Calcium fluorescence imaging provides a means of measuring this activity across large networks on the order of 1,000 neurons [2]. When the excitatory inputs to a particular neuron exceed a threshold voltage, the neuron fires and generates a time-series signal. Because connected neurons generate correlated time-series signals, the likelihood of connection between each pair of neurons is studied. Information theoretic metrics are common ways to measure the connection probabilities. A state-of-the-art measure for measuring excitatory connectivity in calcium imaging neural network data is generalized transfer entropy [3]. Other standard information theoretic measures such as information gain can be used to measure the flow of information in the neural network. After constructing a connectivity matrix whose entries are the pairwise connection scores calculated by a given information theoretic measure, the connectivity matrix is trimmed by a global threshold.

A clear problem with this approach is the scale that leads to impractical computation times. To solve it, we propose a fast method to predict the connectivity of neural networks. We add two ingredients to the classical approach: matrix completion and local thresholding. Instead of calculating all the entries in the connectivity matrix, we only calculate a certain percentage using information theoretic measures. We then use matrix completion techniques to fill the rest of the connectivity matrix. Moreover, instead of using a global threshold, we set a local threshold for each neuron. Our experiments show that this simple method provides faster and better results than those of the classical approaches.

2. BACKGROUND

In this section, we cover the background material for the rest of the paper, including basics of neural networks and previous methods to predict the connectivity of neural networks.

Neural networks. A neural network is a physiological structure in the brain consisting of a group of neurons and their pairwise attachments. These synapses allow the transfer of electro-chemical activity from one neuron to the next [4]. The connections in a neural network are causally directed in the sense that the activity of one neuron will affect the future activity of another neuron it is connected to. The mapping of neural circuitry reveals the building blocks of neural compu-

The authors gratefully acknowledge support from the NSF through awards 1130616 and 1017278, and CMU Carnegie Institute of Technology Infrastructure Award.

¹Neural cultures are groups of neurons grown in a controlled environment which form networks that can exhibit spontaneous behavior.

tation, which provides an understanding for how people learn and reason.

The activity of a neuron is defined by the changes in its membrane potential. A neuron will fire when the level of excitatory input from its neighbors exceeds a particular threshold voltage. The firing of the neuron results in the propagation of an action potential down the axon to synapses, where this signal is passed to other neighboring neurons. The membrane potential of each neuron is measured throughout the course of some network activity, resulting in time-series signals for each neuron in the neural network. These time-series signals are analyzed to reconstruct the neural network’s excitatory connectivity.

Previous methods. A classical framework to predict the connectivity of neural networks includes two components, connectivity matrix construction and global thresholding, as in Figure 1. In the block of connectivity matrix construction, the scores for the likelihood of each pairwise connection are calculated and stored in a connectivity matrix. Then, a global threshold corresponding to the expected number of connections in the neural network is set to decide whether each connection exists. In other words, if the pairwise connectivity score is higher than the global threshold, the corresponding nodes are connected, and vice versa.

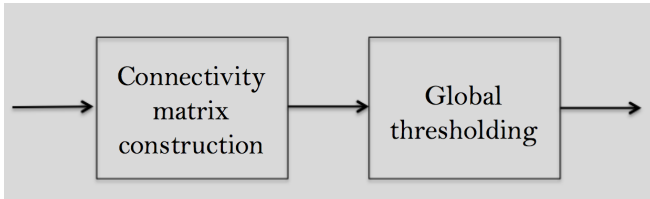


Fig. 1: Classical framework.

A connectivity matrix is a matrix whose entries are the pairwise connectivity scores. Consider a dataset that has N neurons, and form a connectivity matrix $X \in \mathbb{R}^{N \times N}$, whose entry $X_{i,j} = d(\mathbf{s}_i, \mathbf{s}_j)$ denotes the connectivity score from the i th to the j th neuron, where \mathbf{s}_i and \mathbf{s}_j are the time-series signals corresponding to the i th and j th neurons, respectively, and d is a predefined function to measure the connectivity. For greater values of $X_{i,j}$, the likelihood of connectivity is higher.

Two approaches that predict excitatory connections well in the simulated calcium fluorescence data are generalized transfer entropy and information gain. Each of these measures calculates scores indicating the likelihood of a directed excitatory connection between each pair of neurons. Information gain calculates the decrease in uncertainty (or the increase in information) of a neuron’s activity given the activity of another neuron. If the information gain from \mathbf{s}_i to \mathbf{s}_j , denoted $IG(\mathbf{s}_i, \mathbf{s}_j)$, is relatively high with respect to other directed connections in a neural network, it is reasonable that an excitatory connection between the neurons accounts for

the information flow between them.

Transfer entropy [5] measures how useful the causal history of one time-series signal is in predicting the next state of another time-series signal. Rather than standard transfer entropy, we use a modified form to improve results on calcium fluorescence neural network data introduced by [3], called generalized transfer entropy. The generalized transfer entropy from \mathbf{s}_i to \mathbf{s}_j , denoted $GTE(\mathbf{s}_i, \mathbf{s}_j)$, is calculated by ignoring network samples that occur during periods of high global fluorescence activity. Global fluorescence levels above a critical fluorescence threshold indicate burst phases in the neural culture’s activity, which are not representative of the underlying network. For this reason, samples determined to be in these burst phases are not used in the measure. Also, transfer entropy is generalized to calcium fluorescence neural network data by including same time-bin interactions. This means that predictive information about a neuron’s fluorescence activity level can be gathered from the same network sample. This is necessary because neurons interact on a 1ms time basis, while the sampling period of imaging modalities used for calcium fluorescence imaging is roughly ten times as long. Note that for the previous two approaches, the fluorescence data is discretized into bins to calculate the probability distributions.

3. PROPOSED METHOD

The main disadvantage of the previous methods is the computational cost. To solve this, we propose a fast solution by adding two novel ingredients to the classical framework: matrix completion and local thresholding.

Framework. We propose a framework to predict the connectivity of neural networks that includes three components: connectivity matrix semi-construction, connectivity matrix completion, and local thresholding, as in Figure 2. For connectivity matrix semi-construction, we randomly sample a few indices in the connectivity matrix, and compute the corresponding connectivity scores by using the methods discussed in Section 2. For connectivity matrix completion, we fill in the rest of the connectivity matrix by using the matrix completion techniques. For local thresholding, we set a local threshold for each node to decide whether a connection exists.

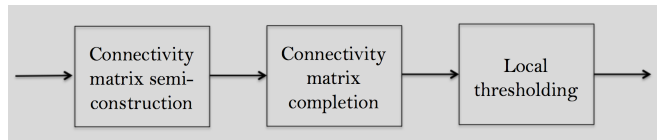


Fig. 2: Proposed framework.

Connectivity Matrix Semi-Construction. Similarly to connectivity matrix construction in the classical framework, connectivity matrix semi-construction calculates the pairwise connectivity scores by using one of the information theoretic approaches. The difference is that connectivity matrix semi-construction computes only a fraction of the connectivity matrix. We randomly calculate some entries in a connectivity

matrix X , denoted as $X_{\mathcal{M}}$, by using one of the information theoretic measures discussed in Section 2.

Connectivity Matrix Completion. Following connectivity matrix semi-construction, we fill in the rest of the connectivity matrix using matrix completion techniques. We use the fact that neurons often connect to other neurons in similar ways. For example, two different neurons may connect to the same neuron. This similarity of connectivity causes a low-rank connectivity matrix. We can then use a low-rank matrix to approximate the entire connectivity matrix, however, minimizing the rank with constraints is known as a non-deterministic polynomial-time hard problem [6]. To solve this, a convex relaxation is achieved by using the nuclear norm [6]. The unmeasured part is then estimated as follows:

$$\hat{X} = \arg \min_Z \|Z_{\mathcal{M}} - X_{\mathcal{M}}\|_2^2 + \lambda \|Z\|_*, \quad (1)$$

where $X_{\mathcal{M}}$ is the calculated entries in the connectivity matrix, $Z_{\mathcal{M}}$ is the corresponding entries of the connectivity matrix approximation Z , λ is the tuning parameter to control the rank, and $\|\cdot\|_*$ denotes the nuclear norm, which is the sum of all the singular values. Since (1) is a convex optimization problem, it can be solved efficiently by any convex optimization solver. Note that solving (1) is much cheaper than calculating the pairwise connectivity scores by using the information theoretic measures. For example, if we calculate only half of the connectivity matrix, we reduce the computation by half.

Local Thresholding. Rather than directly applying a global threshold to the completed connectivity matrix \hat{X} to determine the underlying connections of the neural network, we use a local thresholding method that involves normalizing each row of the connectivity matrix before applying the global threshold. The use of the local threshold is motivated by the fact that each row of the connectivity matrix corresponds to a neuron’s outgoing neighbor connectivity scores.

A neuron in a network reconstructed using global thresholding, may incorrectly have fewer connections than it has in the ground truth network, because the neuron’s connectivity scores (the entries in the row corresponding to that neuron) fall below the global threshold. Remember that before calculating information theoretic measures, the discretization of the fluorescence data is required. Due to the unfavorable boundary conditions in the discretization, some neurons can have smaller connectivity scores than they should have, leading to fewer connections. Similarly, some neurons can have larger connectivity scores than they should have, leading to more connections. In the network, however, neurons often have a similar number of connecting neurons [3]. It makes sense to apply a local threshold for each individual neuron. This is achieved by normalizing each row of the completed connectivity matrix by the l^2 norm,

$$X = \left[\frac{\hat{X}_{i,j}}{\sqrt{\sum_j \hat{X}_{i,j}^2}} \right]_{i,j=0,1,\dots,N}$$

We found that the l^1 norm had similar performance to the l^2 norm. Once the completed connectivity matrix has been

normalized, a local threshold is selected according to the expected number of connections in the neural network.

4. EXPERIMENTAL RESULTS

In this section, we validate the proposed framework on a simulated calcium fluorescence dataset provided by [3]. This dataset has been the subject of the Kaggle connectomics challenge in which participants determine the connectivity of a 1,000 neuron network by using simulated calcium fluorescence data. Participants are encouraged to outperform the standard measures in neural network reconstruction such as generalized transfer entropy and information gain.

Dataset. A simulated dataset is necessary to evaluate reconstruction approaches due to the unavailability of ground truth of connectivity for physical neural networks on the scale of 1,000 neurons. The simulated calcium fluorescence dataset we use to evaluate our reconstruction method is based on dissociated cortical neural cultures with blocked inhibitory GABAergic transmission. This means that all neural activity is based on the excitatory connectivity [3]. A complete description of how the simulated calcium fluorescence data was generated can be found in [3].

A variety of 1,000 neuron datasets with varying levels of clustering and noise are available through the Kaggle connectomics challenge. We compare our results on the four datasets where clustering and noise level are typical of a 1,000 dissociated cortical neuron culture. We refer to these datasets as normal-1, normal-2, normal-3, and normal-4. Each dataset is 1 hour of simulated calcium fluorescence data with a 20ms sampling period. Since each time-series contains 179,500 samples, joint probability distribution calculation for each pair of time-series is computationally expensive. This motivates the use of matrix completion to reduce computation.

Experimental Setup. For each 1,000-neuron dataset, we perform semi-construction at 10%, 20%, 50%, 90%, and 100% of the connectivity matrix entries for both measures, generalized transfer entropy (GTE) and information gain (IG). For IG, an impurity measure must be specified to calculate uncertainty in the time-series signals. We choose Shannon entropy as a standard measure of uncertainty. For matrix completion, we choose the tuning parameter λ to be 0.001. In each experiment, we perform matrix semi-construction with 10 different randomly selected sets of indices. We report the average performance in AUC (area under the ROC curve), for each experiment. These experiments were run on a laptop with 2.50GHz Intel Core i5 processor and 8 GB of RAM.

Results. Figures 3 and 4 compare the performance of our proposed method with GTE and IG respectively at different levels of semi-construction on each dataset. We measure our performance in area under the ROC curve (AUC). The blue (left) and red (right) bars show performance with local thresholding and global thresholding respectively. We see that local thresholding increases reconstruction performance in all cases. Particularly, the performance of the classical approach

Method	Classical approach	Percent semi-construction			
		10%	20%	50%	90%
GTE (hrs)	18.43	1.87	3.71	9.23	16.60
IG (hrs)	6.56	0.68	1.33	3.30	6.01

Table 1: Computational time of the proposed method compared with the classical approach (100% construction).

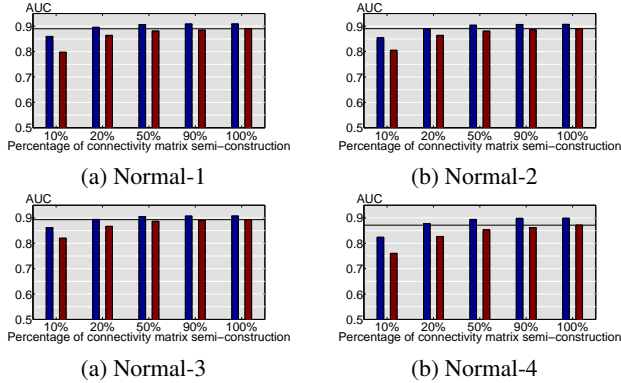


Fig. 3: Performance of the proposed method with GTE with different amounts of matrix semi-construction on normal-1 through normal-4. The horizontal line corresponds to the performance of the classical approach with GTE.

is improved by local thresholding alone.

Our proposed framework with GTE eclipses the performance of the classical approach on all datasets with less than 20% matrix semi-construction. Using our proposed framework with IG, we eclipse the performance of the classical method on all datasets with less than 50% semi-construction. Table 1 shows the average computational time of the experiments. Since construction of the full connectivity matrix takes several hours for both IG and GTE, and matrix completion takes less than a minute for any level of semi-construction, there is an almost linear correlation between saved computational time and the level of matrix completion.

From the results, we see that matrix completion and local thresholding implemented with IG, was able to achieve better performance than the classical approach with half of the computation. Our proposed method implemented with a measure tailored for calcium fluorescence neural network data, GTE, beat the classical method with only 20% of the computation.

Our proposed framework is useful for extracting a high quality network representation from large calcium fluorescence neural network data. The larger the dataset, the more appealing it is to use matrix completion to save computational time, because the required number of sampled entries for matrix completion is $O(N^{1.2} \log N)$, less than $O(N^2)$ [6].

5. CONCLUSIONS

We proposed an efficient algorithm to predict connectivity in neural networks represented by simulated calcium fluorescence data. Instead of computing a full connectivity matrix,

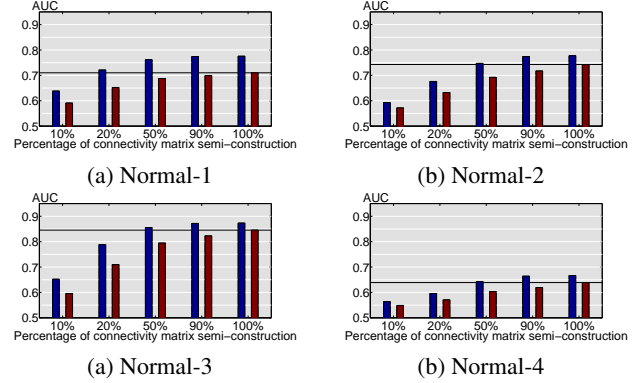


Fig. 4: Performance of the proposed method with IG with different amounts of matrix semi-construction on normal-1 through normal-4. The horizontal line corresponds to the performance of the classical approach with IG.

the proposed method uses only a fraction of this computation and achieves better performance. By computing a subset of the total entries in the connectivity matrix, we use matrix completion to determine the rest of the connection likelihoods. A local threshold is applied to identify the existence of connections. We validate the proposed method on a simulated calcium fluorescence dataset. The proposed method is better than the classical method with 50% computational time using IG and 20% computational time using GTE.

6. REFERENCES

- [1] C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and P. Scheltens, “Small-world networks and functional connectivity in Alzheimer’s disease,” *Cereb Cortex.*, vol. 17, no. 1, pp. 92–99, Jan. 2007.
- [2] J. Soriano, M. R. Martinez, and E. Moses T. Tlusty, “Development of input connectinos in neural cultures,” *PNAS*, vol. 105, no. 37, pp. 13758–13763, May 2008.
- [3] O. Stetter, D. Battaglia, J. Soriano, and T. Geisel, “Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals,” *PLoS Comput Biol.*, vol. 8, pp. 125, Aug. 2012.
- [4] G. E. Palade and S. L. Palay, “The fine structure of neurons,” *J Biophys Biochem Cytol*, vol. 1, no. 1, pp. 69–88, Jan. 1955.
- [5] T. Schreiber, “Measuring information transfer,” *Phys. Rev. Lett.*, vol. 85, pp. 461–464, July 2000.
- [6] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Journal Foundations of Computational Mathematics*, vol. 9, no. 2, pp. 717–772, Dec. 2009.