Reconstruction of Excitatory Neuronal Connectivity via Metric Score Pooling and Regularization

Chenyang Tao CYTAO@FUDAN.EDU.CN
Wei Lin WLIN@FUDAN.EDU.CN
Centre for Computational Systems Biology and School of Mathematical Sciences, Fudan University, Shanghai, China
Jianfeng Feng JIANFENG64@GMAIL.COM
Department of Computer Science, The University of Warwick, Coventry, UK

Editor: Somebody

Abstract

Unravelling the causal link of neuronal pairs has considerable impacts in neuroscience, yet it still remains a major challenge. Recent investigations in the literature show that the Generalized Transfer Entropy (GTE), derived from information theory, has a great capability of reconstructing the underlying connectomics. In this work, we first generalize the GTE to a measure called Csiszár's Transfer Entropy (CTE). With a proper choice of the convex function, the CTE outperforms the GTE in connectomic reconstruction, especially in the synchronized bursting regime where the GTE was reported to have poor sensitivity. Akin to the ensemble learning approach, we then pool various measures to achieve cutting edge neuronal network connectomic reconstruction performance. As a final step emphasize the importance of introducing regularization schemes in the network reconstruction.

Keywords: Csiszár's Transfer Entropy, Metric Score Pooling, Network Regularization, Inverse Correlation

1. Introduction

Understanding the structure and mechanism of the human brain at the cellular and subcellular levels has long been the most challenging issue of science, as echoed in both the recent USA BRAIN project and the EU HBP project. Such a deep understanding will reveal the functions of brain and further inspire the development of the diagnosis, treatment and prognoses of major neurological disorders, such as Alzheimer's disease. We note that recent investigations usually start from understanding learning capability - one of the prominent features of the brain. It is therefore a key issue to reliably recover both the exact wiring patten and the wiring strength of the network at the neuronal level; these are tightly associated with the learning capability of the brain, as the result of the Hebbian learning rule and spike time-dependent plasticities.

Although the traditional neuroanatomic method of axonal tracing can characterize the connectivity for some very small networks, it cannot be applied directly to networks with large scales. Recent advances in calcium imaging has provided an alternative for unveiling the complex neuronal circuitry (Grienberger and Konnerth, 2012). Optical imaging of neuronal activity makes it possible to monitor the simultaneous activity of tens of thousands

of neurons, with a time resolution of 20 ms. With the help of computational algorithms, the causal relationship between neuronal pairs can be determined and the corresponding large scale of the neuronal network reconstructed (Stetter et al., 2012).

To advance research on neuron network reconstruction from Calcium fluorescence imaging data, a platform calling participants to compare and improve their network reconstruction algorithms was established by the committee of 2014 Connectomics Challenge. Synthetic calcium fluorescence recordings generated from realistically simulated neuronal network were presented to the participants to reconstruct synaptic wiring. A few samples, with ground truth topology, were provided to train participants' models, one validation set without ground truth topology was provided to validate their solutions, and the performances of the solutions were benchmarked on a test sample using the so-called Area Under ROC Curve (AUC) score. In this short paper, we introduce our approach for solving the challenge, which finally ranks 9^{th} on the platform.

The remainder of this paper is organized as follows. In Section 2, we first describe the preprocessing steps adopted, then we detail the CTE measure, score pooling and regularization procedure. The results are presented in Section 3. Finally in Section 4, we discuss the limitations of our approach and point out a few potential future directions.

2. Methods

2.1. Preprocessing of Calcium imaging

The following preprocessing steps have been adopted to generate input data for computing the metric scores used in the reconstruction of neuronal wiring.

Two schemes have been used to separate the synchronized and unsynchronized dynamical regimes. First one is simple thresholding, the period during which mean Calcium imaging intensity exceeding certain threshold is identified as synchronized bursting regime. Multiple thresholding parameters are used (from 0.12 to 0.25). The second approach explicitly extracts the synchronized dynamics and deflates it from the individual recordings. Specifically, the first eigenvector of the principal component of the raw fluorescence data is identified as synchronized dynamics and is projected out from the recordings.

Both the simple discretization and more elaborate OOPSI package (Vogelstein et al., 2010) were used to infer the spike trains from the Calcium waves. Signals with and without deflation of the synchronized dynamics were all discretized using the above two schemes. For the OOPSI scheme, we used the fast_oopsi implementation to speed up the preprocessing. The iteration runs were set to 5-8 depending on the SNR of the data. After filtering with OOPSI, the 1% largest non-zero entries were identified as spiking time points while the rest were identified as noise and discarded.

We also separated the individual responses during synchronized bursting. We first identified the spiking time points of the synchronized dynamics using an OOPSI filter. Then the response of individual neurons, during the synchronized bursting period, was characterized as the Calcium imaging intensity increase at the spiking time point.

2.2. Csiszár's Transfer Entropy

In probability theory, a divergence measure is a function $D(P \parallel Q)$ that measures the difference between two probability distributions P and Q. The most widely used divergence measure is the Kullback-Leibler (K-L) divergence, with the mutual information as a special case. This idea was later generalized by Csiszár, which resulted into a family of divergence measures (Csiszár, 1963). This is known as the Csiszár's f-divergence, which is defined as

$$D_f(P \parallel Q) \equiv \int_\Omega f\left(\frac{dP}{dQ}\right) dQ.$$

where f is an convex function satisfying $f(1) = 0$.

The transfer entropy (TE) is a non-parametric statistic measuring the amount of directed (time-asymmetric) transfer of information between two random processes (Schreiber, 2000). It could be interpreted as the reduced uncertainty of future X given the present Y, or the K-L divergence of the transition probability with or without the knowledge of Y. Replacing the log function in TE with the convex function f in the Csiszár's f-divergence, we obtain the more general Csiszár's Transfer Entropy by analogy:

$$\mathrm{CTE}_{Y\to X} = \int_\Omega f\left(\frac{dP(X_{t+1}|X_t^{(k)})}{dP(X_{t+1}|X_t^{(k)}, Y_t^{(k)}, Y_{t+1})}\right) dP(X_{t+1}, X_t^{(k)}, Y_t^{(k)}, Y_{t+1}).$$

Here each $Z_t^{(k)}$ denotes the delay-embedded state vector $(Z_t, \cdots, Z_{t-k+1})$, and the additional Y_{t+1} in the conditioning conforms to the GTE used in Stetter et al. (2012). Binary valued spike trains were used to calculated CTE/GTE. In this study, we use the α-divergence (Liese and Vajda, 2006) specified by

$$f(t) = \begin{cases} \frac{4}{1-\alpha^2}\left[1 - t^{(1+\alpha)/2}\right], & \text{if } \alpha \neq \pm 1, \\ t \ln t, & \text{if } \alpha = 1, \\ -\ln t, & \text{if } \alpha = -1. \end{cases}$$

As the K-L divergence is a special case of the *alpha*-divergence, so their performance could be directly compared. Here, the ideal value of α should maximize the AUC score in the training sample. In this studied we discretized the data into binary code indicating whether the neuron is firing, thus making it comparable to GTE. We note more refined binning of the neuron's firing intensity will improve the performance at the cost of larger memory usage. Some other convex functions have also been tested and produce similar best performances (data not shown).

2.3. Correlation metrics

The conventional Pearson's correlation was also calculated to generate the pooled statistics for optimal connectivity reconstruction as it could be obtained cheaply and proved to be a quite good metric score when the data is properly preprocessed. Specifically, we used the correlation and delayed correlation (with lag 1). The correlation metrics were calculated from the following input data: individual response during the synchronized bursting period, OOPSI-filtered spikes during unsynchronized bursting period. The spikes used to calculate

the correlation metrics are real valued to reflect the spiking intensity during bursting and it is more informative compared to correlation calculated from binary valued spike trains. We also tested the performance of more general but computationally more intensive nonlinear kernel-based correlation metric (Bach and Jordan, 2003) after the challenge and is briefly discussed in **Supplementary Information**.

2.4. Pooling of different metric scores

Two simple approaches were used to integrate the evidence from different metrics and different preprocessing schemes using the training data. Specifically, we considered the Bayesian posterior probability and a linear combination of metrics. First, the original score obtained from different metric or preprocessing schemes are normalized to the interval $[0, 1]$ according to their ranks. Then the bayesian posterior probability for the corresponding link being true, given the observed (normalized) metric score $R^{data}_{X\to Y}$, is calculated by

$$P(S_{X\to Y} = 1|R^{data}_{X\to Y}) = \frac{P(S_{X\to Y} = 1, R^{train}_{X\to Y} = R^{data}_{X\to Y})}{P(R^{train}_{X\to Y} = R^{data}_{X\to Y})}$$

where $S_{X\to Y}$ represents whether there is a true link from X to Y and $R_{X\to Y}$ is a vector of normalized scores. The probability in the above formula could be estimated either by kernel smoothing or binning. To ensure sufficient samples for estimating the probability, we restricted the dimension of $RC_{X\to Y}$ to two. We also use the following simple linear combination to aggregate the evidence from two different metrics $\tilde{R}$ and $\hat{R}$:

$$R^{joint}_{X\to Y} = \omega\tilde{R}^{data}_{X\to Y} + (1-\omega)\hat{R}^{data}_{X\to Y},$$
$$\text{where } \omega = \arg\max_{\omega} \text{AUC}(\omega\tilde{R}^{train}_{X\to Y} + (1-\omega)\hat{R}^{train}_{X\to Y}).$$

These two approaches defining the basic hybridization operation on the pool of all metric scores. Enlightened by staked ensemble learners (Zhou, 2012), we adopted an evolutionary-like hybridization procedure that heuristically mates two relevant[1] or best performing metric scores and then adds their best offspring to the pool. We then repeated this procedure until the best AUC score in the pool no longer increased.

2.5. Regularization on the recovered network

We observe that for all the metrics scores we obtained, the degree distribution of the reconstructed networks differs from the genuine wiring that generated the data. The presynaptic and postsynaptic links of the estimated hub nodes are often overestimated while some of the non-hub nodes are disconnected from the estimated network. In this light we argue that in order to obtain more realistic reconstruction, we must regularize the network topology - to some extent. In this study, we did this by explicitly reweighting the score metrics to suppress the links related to the hub nodes and to encourage the links that wire the disconnected nodes back to the network. The reweighting procedure is outlined in Algorithm 1 in **Supplementary Information**.

1. Relevant in the sense that they are derived from same input data or same metric score.

2.6. Evaluation of the reconstruction performance

The network reconstruction was considered as a binary classification problem. The solution returns a numerical score for each directed neuron pair indicating the confidence that there is a directed connection, with higher values indicating a more likely connection. The results of the classification, obtained by thresholding the prediction score, may be represented in a confusion matrix, where **tp** (true positive), **fn** (false negative), **tn** (true negative) and **fp** (false positive) represent the number of examples falling into each possible outcome. The sensitivity (also called true positive rate or hit rate) and the specificity (true negative rate) as:

True positive ratio = tp/pos
False positive ratio = fp/neg

Here **pos** = **tp**+**fp**, **neg** = **tn**+**fn** indicating the total number of connected and unconnected pairs. The prediction results are evaluated with the AUC, which corresponds to the area under the curve obtained by plotting the "True positive ratio" against the "False positive ratio" by varying a threshold on the prediction values to determine the classification result.

3. Results

In this section, we present an empirical study of our proposed procedure on the four training sets (**normal-**1 $\sim$ 4) provided in Connectomics Challenge. Each of these training sets is comprised of approximately $170,000$ continuous recordings sampled at 50 Hz of $1,000$ neurons with 1.2% connected pairs. Interested readers may refer to (Stetter et al., 2012) for details of the simulation setup.

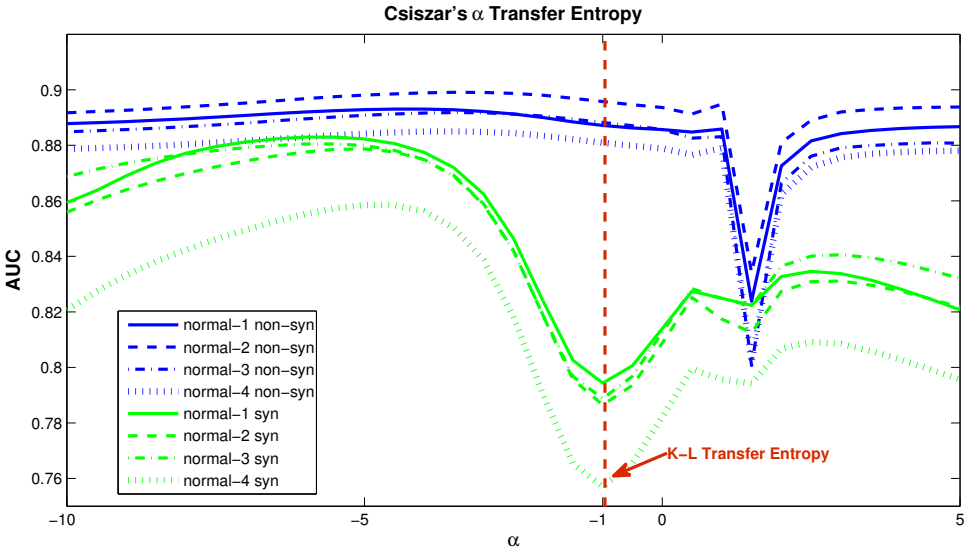
3.1. CTE

We compared the performance of the CTE with the GTE. A family of the CTE was obtained by varying the parameter α and the resulting α-AUC curves are shown in Figure 1. For the 4 datasets, the peak of the α-AUC curves consistently appeared around $\alpha \approx 4$. And the differential sensitivity with respect to the dynamical regime is surprising. While the CTE only offers small advantage in the non-synchronized bursting regime ($\sim 4 \times 10^{-3}$), it significantly improved the reconstruction in the synchronized bursting regime ($\sim 9 \times 10^{-2}$ in this illustration, and even more drastically using some other discretization scheme). It is interesting to notice that the AUC score of the traditional K-L Transfer Entropy happens to fall on the bottom of the valley in the synchronized regime.

3.2. Pooling metrics scores

The effectiveness of pooling different metric scores is presented in Table 1. In the upper panel of the table AUC scores of a representative subset of the the raw metric scores used along with their pooled metric score (using all the raw metric scores) are tabulated for the 4 validation datasets. The tuning parameters are optimized for dataset 'normal-3' and used by the rest of datasets. As shown, pooling significantly boosts the performance of the reconstruction evaluated by AUC scores. The fifth column of lower panel of the table gives

Figure 1: The AUC score of Csiszár's Transfer Entropy with different values of the parameter α. Blue curves are obtained from the non-synchronized regime while the green curves from the synchronized regime. The intersections with the red curve correspond to the traditional K-L GTE.



one of the best (and fastest) winning solutions based on inverse correlation[2], and combining it with our best solution (column **POOL**) gives extra $4 \sim 6 \times 10^{-3}$ boost in AUC score (column **BEST**), which beats all best solutions during the contest. Detailed description of those metric scores and whether they are used in the challenge could be find in Table 2.

Table 1: AUC Score with Pooling and Regularization

	CORR1	**CORR1D**	**CORR2**	**CORR2D**	**GTE1**	**GTE2**	**POOL**
normal-1	0.8888	0.8585	0.8918	0.5568	0.8892	0.6503	0.9240
normal-2	0.8934	0.8556	0.8894	0.5328	0.8930	0.6289	0.9256
normal-3	0.8906	0.8542	0.8920	0.5611	0.8932	0.6545	0.9248
normal-4	0.8876	0.8499	0.8844	0.5380	0.8721	0.6216	0.9217
	CORR2	**CORR2i**	**CORR2o**	**CORR2io**	**MIC**	**MICo**	**BEST**
normal-1	0.8918	0.8955	0.8997	0.9001	0.9412	0.9422	0.9465
normal-2	0.8894	0.8938	0.8986	0.8991	0.9412	0.9422	0.9473
normal-3	0.8920	0.8966	0.9009	0.9015	0.9394	0.9405	0.9461
normal-4	0.8844	0.8883	0.8928	0.8934	0.9376	0.9385	0.9441

† **D**: delayed correlation, **i**: regularizing on input (postsynaptic connections), **o**:regularizing on output (presynaptic connections), **io**:regularizing both input and output)

Table 2: Description of the metric scores

	CORR1	**CORR2**	**GTE1**	**GTE2**	**POOL**	**BEST**
individual response (syn)	O	-	-	-	-	-
OOPSI-filtered (non-syn)	-	O	-	-	-	-
discretization (syn)	-	-	O	-	-	-
discretization (non-syn)	-	-	-	O	-	-
use entire sequence	-	-	-	-	-	O
binary	-	-	O	O	-	-
used in challenge	O	O	O	O	O	-
challenge submission	-	-	-	-	O	-

3.3. Network regularization

The gain using network regularization is presented in the lower panel of Table 1. The column name indicates the regularization used. For the raw metric score **CORR2** this could bring up the AUC score by about 1×10^{-2}. For the inverse correlation metric it can still elevate the AUC score by 1×10^{-3}. We noticed that regularizing the output links resulted in larger gain compared with regularizing the input links. We note that our regularization scheme is definitely not a universal fix as it certainly depends on the assumption that the distribution

2. http://www.kaggle.com/c/connectomics/forums/t/8186/fast-matlab-code-to-get-a-score-of-93985

of training data and testing data is the same, and violation of this assumption will lead to deteriorated performance.

3.4. Challenge results

The final performance of our challenge solution and post-challenge solution in the 2014 Connectomics Challenge is presented in Table 3 together with the winner's performance. Our best post-challenge solution outperform the best challenge solution by a large margin.

Table 3: Result Table

team name	**killertom**
final private leaderboard performance	0.93011 (ranking 9^{th})
winner's performance	0.94161 (team **AAAGV**)
our post-challenge best performance	**0.94663** (**BEST** in Table 1)

4. Discussion

Our method is based on linear combination of correlation coefficient and CTE, using data preprocessed with simple discretization, OPPSI filter and PCA. Most of the winning teams' solutions relied on correlation-based metrics, and inverse correlation in particular. Some teams used more sophisticated machine learning tools such as deep CNN, random forest, SVM, etc. Some participants also came up with certain network regularization schemes such as network deconvolution or directly including node-wise relative metric score into the model. Most teams emphasized the paramount importance of preprocessing the noisy calcium imaging data. Our approach seems to be the only one which still extensively uses entropy-based statistics among all the winning parties, possibly due to the costly computational burden involved. We resolve this by optimizing the MATLAB subroutine provided by the organizer which ended up running 20 times faster on desktop than the C++ implementation also provided by the organizer on cluster. As shown in column **BEST** in Table 1, there is still much room for improvement by combining our approach with other winning teams' solutions, even if their AUC score is 1×10^{-2} better than our results.

One significant problem that applies to most of the winning team approaches, including ours, is that the optimal predictability for connectivity is 'learnt from training samples' rather than 'inferred from the dynamics observed', as in reality it is infeasible to obtain real training samples and simulation based surrogates might be biased. We argue that it is the dynamical properties that matter, and instead of statistical solutions, we should start looking for apparatus from the theory of dynamical systems (Sugihara et al., 2012). Also, most participants are still determining the causal links in a pair-wise fashion, with the possibility of gaining information from a more holistic perspective is left uncharted. Those computationally feasible non-linear association measures (Gretton et al., 2008) might serve as substitutes for those computationally demanding entropy-based statistics.

Acknowledgments

First, we thank the challenge organizers for their hard work setting up such an excellent challenge. We also would like to thank Prof. David Waxman and Dr. Wenbo Sheng for the fruitful discussions. C Tao is supported by China Scholarship Council. W Lin is supported by NNSF of China (Grant Nos. 11322111 and 61273014), and from Talents Programs (Nos. 10SG02 and NCET-11-0109). J Feng is s a Royal Society Wolfson Research Merit Award holder, partially supported by National Centre for Mathematics and Interdisciplinary Sciences (NCMIS) in Chinese Academy of Sciences.

References

Francis R Bach and Michael I Jordan. Kernel independent component analysis. *The Journal of Machine Learning Research*, 3:1–48, 2003.

I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad*, 8:85–108, 1963.

Arthur Gretton, Kenji Fukumizu, Choon Hui Teo, Le Song, Bernhard Schölkopf, and Alex J Smola. A kernel statistical test of independence. 2008.

Christine Grienberger and Arthur Konnerth. Imaging calcium in neurons. *Neuron*, 73(5): 862–885, 2012.

Friedrich Liese and Igor Vajda. On divergences and informations in statistics and information theory. *Information Theory, IEEE Transactions on*, 52(10):4394–4412, 2006.

Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

Olav Stetter, Demian Battaglia, Jordi Soriano, and Theo Geisel. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS computational biology*, 8(8):e1002653, 2012.

George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.

Joshua T Vogelstein, Adam M Packer, Timothy A Machado, Tanya Sippy, Baktash Babadi, Rafael Yuste, and Liam Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of neurophysiology*, 104(6):3691–3704, 2010.

Zhi-Hua Zhou. *Ensemble methods: foundations and algorithms*. CRC Press, 2012.

Supplementary Information

Algorithm 1: Regularizing the network via reweighing

1. Sort the neurons according to their largest in(out) score
   $RC_{i,\cdot}$: the column corresponding to i^{th} largest in $\{\max(C_{k,\cdot})\}$
2. Reweigh the first K in(out) links for each neuron via
   $RC^{v_1}_{i,j} = (1 + \alpha\frac{j}{n})RC_{i,j}$
3. Calculate the probability of the in(out) links of the i^{th}-ranking neuron being connected given the training data (sorted in the same fashion)
   $PS_{i,\cdot} = \underset{|k-i|<B}{\text{average}}(RS^{train}_{k,\cdot})$
4. Smooth individual neuron's PS score
   $PS^{sm}_{i,\cdot} = \text{smooth}(PS_{i,\cdot})$
5. Prioritizing the entries with PS^{sm} exceeding the threshold γ while taking the current estimate of connectivity strength into consideration
   $RC^{v_2}_{i,j} = \chi_{[PS^{sm}_{i,j}>\gamma]}(1 + PS^{sm}_{i,j} + \beta RC^{v_1}_{i,j}) + \chi_{[PS^{sm}_{i,j}\leq\gamma]}RC^{v_1}_{i,j}$
6. Enforcing a minimum number of L in(out) links for each neuron
   $RC^{v_3}_{i,j} = \chi_{[j<l]} + RC^{v_2}_{i,j}$

The set of tuning parameters $\{K, L, B, \alpha, \beta, \gamma\}$ are selected to maximize AUC score in the training data.

kernel-based correlation metric

We used the generalized-variance in Jordan's ICA paper to characterize the nonlinear kernel correlation. Specifically, only the first eigenvalue is used. The kernel-based correlation gives similar performance in instantaneous coupling (Pearson 0.888 $\approx$ 0.896 V.S. kernel 0.889 $\approx$ 0.896) while it significantly outperforms Pearson's correlation in lag-1 correlation (Pearson 0.550 $\approx$ 0.564 V.S. kernel 0.706 $\approx$ 0.711). This is because the lag-1 dynamics exhibits a highly nonlinear pattern. More detailed results and discussion will be presented in a separate paper.