Directed Information for Point Process Systems

Shailaja Akella
Allen Institute
Seattle, Washington
shailaja.akella@alleninstitute.org

Andre M. Bastos
Department of Psychological Sciences
Vanderbilt University,
Nashville, Tennessee
andre.bastos@vanderbilt.edu

Jose C. Principe
Department of Electrical and Computer Engineering
University of Florida,
Gainesville, Florida
principe@cnel.ufl.edu

Abstract

Owing to neurotechnological advances in electrode design, it is now possible to simultaneously record spiking activity from hundreds to thousands of neurons. Such extensive data provides an opportunity to study how groups of neurons coordinate to form functional ensembles that ultimately drive behavior. Since the spike train space is devoid of an algebraic structure, quantifying causal relations between the neuronal nodes poses a computational challenge. Here, we combine techniques from information theory and kernel-based spike train representations to construct an estimator of directed information for causal analysis between neural spike train data. Via projection of spiking data into a reproducing kernel Hilbert space, we avoid tedious evaluations of probability distribution while engaging computations in a non-linear space of (possibly) infinite dimensionality. Additionally, the estimator allows for conditioning on 'side' variables to eliminate indirect causal influences in a multi-neuron network. Extensive analyses on a simulated six-neuron network model comprising of different neuron types and causal topologies show that the devised measure identifies directional influences accurately that would be otherwise inaccessible with traditional correlation measures. Finally, we apply the metric to identify direct causal interactions among neurons recorded from cortical columns of visual-area 4 of monkeys performing a delayed match to sample task. Our results reveal an interesting reorganization of neuronal interaction patterns within a cortical column on visual stimulation.

1 Introduction

Due to the mathematical differences between continuous and point-process signals, current analysis of neuronal connectivity is largely limited to bivariate cross-correlation metrics. However, these methods provide little insight into the directional nature of neuronal interactions and usually do not consider the point-process nature of spike-train data. In this paper, we propose a non-parametric directed information (DI) measure for causal analysis in the point-process space. Here, we present the formal definitions of DI.

Directed information: Directed information is an asymmetrical measure of information flowing from one random process to another [1, 6]. Unlike mutual information (MI), the definition of DI explicitly accounts for feedback between two coupled systems. In our notation, DI between two

Information-Theoretic Principles in Cognitive Systems Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

stochastic processes, X and Y, can be written as in eq. (1), where X^n denotes the successive samples of the process, X, i.e., $X^n = [X_1, X_2, ..., X_n]$. In the absence of a causal influence of X on Y, the second term reduces to an entropy estimation, and therefore, $I(X^n \to Y^n) = 0$. DI is also upper bound by MI when communication between two nodes is carried over a purely feedforward channel.

$$I(X^n \to Y^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}X^i) \tag{1}$$

Causal conditional directed information: In multinodal networks, the connection between two nodes may be mediated by other members of the network. We refer to the time series observed on the other nodes as 'side' information. In order to identify direct causal influences in a network, the contributions from 'side' nodes are accounted for by causal conditioning. In this case, the causal conditional DI between two processes X and Y, conditioned on the side process, Z, is defined as,

$$I(X^n \to Y^n||Z^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}Z^i) - H(Y_i|Y^{i-1}X^iZ^i). \tag{2}$$

2 Directed Information Estimator

Proposed estimator: Computation of DI requires prior knowledge about the joint probability distribution function of the random variables. However, access to such information is mostly unknown in neuroscience scenarios. Our proposed estimator circumvents any estimation of the probability distribution functions from data. Instead, we exploit the properties of a Hilbert space approach for entropy estimation using the matrix-based Renyi's α-order entropy proposed in [4]. The estimator defines entropy over the eigenspectrum of the normalized Gram matrix, a Hermitian matrix of the projected data in a reproducing kernel Hilbert space (RKHS). Specifically, the functional definitions fulfill similar axiomatic properties of Renyi's α-order entropy ($\alpha > 0$), which is a generalization over the well-known Shannon's entropy [8]. Given a random variable $X = \{x_1, x_2, ..., x_n\}$ and a normalized Gram matrix A evaluated on X using the real-valued positive definite kernel, $\kappa : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, the matrix-based estimator of Renyi's α-order entropy is given in eq. (3),

$$S_\alpha(A) = \frac{1}{1-\alpha}\log_2\left[\sum_{i=1}^{n}[\lambda_i(A)]^\alpha\right] \tag{3}$$

where $A_{ij} = \frac{1}{n}\frac{K_{ij}}{\sqrt{K_{ii}K_{jj}}}$ and $\lambda_i(A)$ denotes the i^{th} eigenvalue of A and K is the non-normalized Gram matrix. For evaluations of joint entropy, an extension of the matrix-based entropy measure uses Hadamard products to convey the joint representation of two random variables. In this extension, joint entropy between two random variables, X and Y, is defined as in eq. 4, where A and B are normalized Gram matrices evaluated on X and Y, using positive definite kernels, κ_1 and κ_2.

$$S_\alpha(A, B) = S_\alpha\left(\frac{A \circ B}{tr(A \circ B)}\right) \tag{4}$$

Substituting the above entropy estimation into the DI formulation, we can evaluate the information metric as in eq. (5).

$$I_\alpha(X^n \to Y^n) = \sum_{i=1}^{n} S_\alpha(Y_i|Y^{i-1}) - S_\alpha(Y_i|Y^{i-1}, X^i) \tag{5}$$

However, we are still faced with the challenge of defining a positive-definite kernel for point processes. We take support of kernel constructs that introduce basic structure to the point process space [7].

Here, we briefly summarize the approach. Firstly, each point process is brought into the continuous L_2 space as intensity functions via convolutional operators. Next, using an RKHS inducing kernel, κ and evoking the kernel trick, the inner product of two intensity functions x and y in the RKHS is computed by a scalar evaluation of the kernel in the input space, i.e., $\langle x, y \rangle_{\mathcal{H}} = \kappa(x, y)$. In this way, the required normalized Gram matrix, A, is obtained as a symmetric positive definite matrix representing the pairwise inner product of the intensity functions in the Hilbert space.

$$\kappa^r_{Sch}(x, y) = \exp\Big\{ \int_T -\frac{1}{\sigma_\kappa}(x(t) - y(t))^2 dt \Big\} \tag{6}$$

$$x(t) = \sum_i \mathbb{I}(t_i - w \leq t \leq t_i + w) \tag{7}$$

Here, we note an important consequence of DI estimation. Although theoretically, the DI between independent time series is equal to zero, in practice, empirical values are small and non-negative. This is because any DI estimation does not incorporate the true probability distribution of the involved time series. We address this issue via appropriate significance testing for each connectivity (See Appendix A.1).

Parameter selection: Our formulation of DI depends on four hyperparameters: 1) the kernel size (σ_κ), 2) the width of the rectangular intensity operator (w), 3) the order α, and 4) memory of the DI measure (n). The kernel size dictates the geometrical extent of the inner product in the RKHS. It is important to note that for projections into the same RKHS, the kernel size must be the same for all entropy evaluations. The choice of α is associated with the task under study. For emphasis on rare events, for eg. in a discriminative task, α must be chosen close to 1 [4], while higher values of α (> 2) characterize modal behavior. Lastly, the memory, n, of the DI measure must be selected to capture signal transmissions that occur within the timescale of interest. It must be noted that the total time dependency would ultimately be a combined effect of the width, w, of the rectangular intensity operator and the memory, n, of the DI measure. Throughout all analyses, we maintain the parameters at the following values: 1) σ_κ set using Scott's rule of thumb on the source variable, 2) w = 120ms, 3) α = 1.01, and 4) n = 20ms.

3 Results

Simulations Before application to real data, we evaluate the performance of our proposed estimator on synthetically generated spike trains with known causal influences. We tested the model on a six-neuron network model with connections that closely mimic the complex interactions between neurons. All neuronal activity was simulated using the Izhikevich neuron model [5]. The model is governed by the following set of equations (Eqs. 8.5 and 8.6 in [5]),

$$C\dot{v} = k(v - v_r)(v - v_t) - u + I \tag{8}$$

$$\dot{u} = a\{b(v - v_r) - u\} \tag{9}$$

$$\text{if } v \geq v_{peak}, \text{ then } v \leftarrow c,\ u \leftarrow u + d \tag{10}$$

where v is the neuron's membrane potential, and u is the membrane recovery variable. Description of the hyperparameters and their corresponding values for a regular-spiking neuron were maintained as in [5].

Connectivity between the neurons is shown in Figure 1**A2**(left, solid lines) where 5 excitatory $\{n_{1-3}, n_{5-6}\}$ and an inhibitory neuron (n_4) are involved in a complex network designed to include indirect influences (eg. $n_1 \leftrightarrow n_3$), explicit driver neurons (eg. n_5) and multi-input neurons (eg. n_6) Each neuron was simulated using one out of the three types of input current, I_{in} - 1) constant input, 2) uniform input: a uniformly distributed value between $(0 - 150)$ pA, or 3) bursty input: two 150 ms long pulses of 300 pA, each supplied at a random time, interleaved with a uniformly distributed current between $(0 - 150)$ pA. Input type 3 was supplied to create a system of bursty neurons. Specifically, uniform input current was supplied to neurons n_1, n_2, n_3 and n_6. Whereas n_5 received a constant input current of 150 pA, n_4 was simulated to mimic an interneuron such that

it received a bursty input current. The magnitude of excitation current, I_e (or inhibition current I_i) from an excitatory neuron (or inhibitory neuron) in a connection was kept constant throughout each simulation. These values were set at $I_e = 350$ pA and $I_i = 180$ pA. The total current, I, supplied to each neuron comprised of I_{in} and $I_{i(e)}$. Membrane noise was generated as Gaussian noise and combined with the neuron's membrane potential, v, at each time point. The system was simulated at a fixed sampling frequency of 500 Hz. The synaptic delay between any two connected neurons was maintained at 5 ms, and the memory for all DI evaluations was set at 20 ms (or 10 samples). Finally, each simulation was run using time bins of 1 ms, trials lasted for about 1s, and a total of 100 trials were simulated. Figure 1**A1** summarizes our results where the final connection matrix is presented in **A2**. The model is able to identify all direct causal patterns. Although the current system includes many potentially indirect influences, the causal model only identified one edge to be significant, demonstrating high robustness to spurious causal influences. Overall, our model achieved a connectivity detection accuracy of 97.2%.

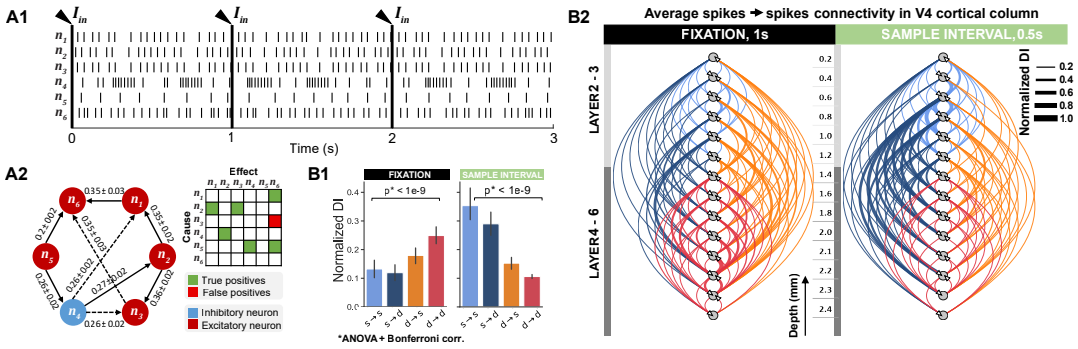

Figure 1: **A. Identification of direct influences in a multi-neuron system**. A1. Example data presenting 3 trials in each neuron. A2, (Left) DI estimates identified by the model. Solid lines-true direct causal directions, dashed lines- identified indirect causal influences. (Top-right) Final estimated adjacency matrix of direct causal influences. **B. Spike–spike interactions**. B1, Connectivity trends between superficial (s, layers 2-3) and deep (d, layers 4-6) layer neurons in the fixation (1s) and early sample (0.5s) intervals. Mean $\pm$ SEM across all direct connectivity. B2, Direct causal influences between neurons in the V4 cortical column across all sessions (N = 123 units) during the same intervals.

Laminar organization of neuronal connectivity In monkeys performing a visual delayed-match to sample task [2], we analyzed the causal interactions between single neurons within individual columns of visual-area 4 (V4) (see Appendix A.2). To determine whether the neuronal interaction patterns change through the trial, we analyzed direct influences in the fixation and early sample interval, [0-0.5]s. In figure 1**B2**, we use directed graphs to summarize the general causal connectivity in the V4 cortical column across subjects for each task interval. Each edge and its direction is identified by aggregating the identified causal connections in individual cortical columns. In contrast, the edge weights describe the strength of the connection as the averaged DI overall direct connections. Across fixation and sample intervals, we observed a reversal in the within-column causal influence patterns in V4 figure 1**B1**. While fixating, the dominant pathway of causal influence included the deep $\rightarrow$ superficial and the deep $\rightarrow$ deep layer connections. Overall, superficial layer neurons were more specialized in processing information during the sample period, while deep layer neurons directed information processing during fixation. Fitting with the hierarchical organization of the visual cortex, these results suggest functional separation between layers of the cortical column. This functional specialization is consistent with the anatomical connection patterns of the primate visual cortex ([3]) wherein feedforward connections to subsequent cortical areas originate mainly from superficial pyramidal cells. In this view, the causal network identified in our analysis may be reflective of stimulus-specific feedforward processing in the superficial layers.

4 Conclusion

Our implementation of directed information for point process spaces presents a data-driven approach to the problem of causal inference in neuroscience. The proposed methodology imposes no assumptions on the signal structure, avoids tedious evaluations of the probability density functions and enables identification of non-linear causal structures between random variables. The method draws support from kernel representation of spike train data wherein non-linear projections into a high-dimensional space allow for data-driven interpretations. Finally, we would like to emphasize that in the context of brain connectivity patterns, it is impossible to distinguish absolute direct influences from indirect influences since not all nodes of the functional network are sampled. While the problem is not theoretical, it has important implications for interpreting neuronal interactions. Under the circumstances, care must be taken when making causal inferences.

References

[1] Pierre-Olivier Amblard and Olivier JJ Michel. "The relation between Granger causality and directed information theory: A review". In: *Entropy* 15.1 (2013), pp. 113–143.

[2] André M Bastos et al. "Layer and rhythm specificity for predictive routing". In: *Proceedings of the National Academy of Sciences* 117.49 (2020), pp. 31459–31469.

[3] Daniel J Felleman and David C Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." In: *Cerebral cortex (New York, NY: 1991)* 1.1 (1991), pp. 1–47.

[4] Luis Gonzalo Sanchez Giraldo, Murali Rao, and Jose C Principe. "Measures of entropy from data using infinitely divisible kernels". In: *IEEE Transactions on Information Theory* 61.1 (2014), pp. 535–548.

[5] Eugene M Izhikevich. *Dynamical systems in neuroscience*. MIT press, 2007.

[6] James Massey et al. "Causality, feedback and directed information". In: *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*. 1990, pp. 303–305.

[7] Il Memming Park et al. "Kernel methods on spike train space for neuroscience: a tutorial". In: *IEEE Signal Processing Magazine* 30.4 (2013), pp. 149–160.

[8] Alfréd Rényi et al. "On measures of entropy and information". In: *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 547-561. Berkeley, California, USA. 1961.

A Appendix

A.1 Experimental Setup

The data analyzed in this study comprised multi-laminar recordings from the visual area 4 (V4) of two macaque monkeys (*Macaca mulatta*) trained to perform a delayed match to sample (DMS) task. Each subject was implanted with linear U and V recording arrays placed over the visual cortex, optimized to enable perpendicular recordings relative to the cortical folding. For each session, the number of laminar probes varied between 1 and 3. Each probe comprised 16 electrodes with an intersite spacing of 100 or 200μm, culminating in a total linear sampling of 3.0 to 3.1 mm on each probe. Channels from the top of the cortex to a depth of 1.2 mm were classified as superficial layer channels, and deep layer channels were those below 1.2 mm. All surgical and animal care procedures were approved by the Massachusetts Institute of Technology (MIT)'s Committee on Animal Care and were conducted following the guidelines of the National Institute of Health and MIT's Department of Comparative Medicine.

The experiments were conducted inside a sound-proof behavioral testing booth comprising a primate chair and an LCD monitor (ASUS, Taiwan). Subjects were trained to perform the DMS task using positive reinforcement. Monkeys were to fixate on a point in the center of the LCD screen for a duration of 1s. If fixation was held for the entire duration, the screen was presented with one of three cue objects. The object remained on the screen for 1s and disappeared during the following delay period. While the delay period for monkey-2 lasted for a uniformly picked time interval between $0.5 - 1.2$s, for monkey-1, the delay period was a fixed time interval of 1s. At the end of the delay period, a search array consisting of the cued item and one or two distractor objects appeared on the LCD screen, each occupying a different visual quadrant. The distractor and its position were

randomly chosen. A successful trial entailed the subject performing a saccade towards the cued object, for which they received a few drops of diluted juice for positive reinforcement. The trial was terminated if the subject broke fixation at any time. Additionally, the predictability of the cued object was manipulated either with repetitive or novel cuing. During novel cuing, the cue was randomly sampled from a set of 3 objects and presented in each trial. Contrarily, repetitive cuing consisted of the same cue presented in each trial. Each block lasted for 50 trials, where the initial cuing-type was randomly chosen.

Throughout each session, neural activity was sampled at 30 kHz and bandpassed between 0.3 Hz and 7.5 kHz using a first-order Butterworth filter. A Plexon offline sorter was used to manually perform spike sorting. For analyses, spiking activity was downsampled to 500Hz.

A.2 Causal inference: Significance testing

In practice, all estimated DI measures (between causally and non-causally connected nodes) are greater than zero. To identify influences that drive significant causal influences, significance testing methods require knowledge of the probability density function of the DI measure under the null hypothesis of non-causality. However, given the statistical variability of neural activity, the null hypothesis distribution is ever-changing and difficult to pin down. In our studies, we test the significance of the DI measures for neuronal interactions using a random permutation procedure to build a baseline null hypothesis distribution. The calculated DI measure from the actual data is then compared against this baseline distribution to assess the significance levels. Specifically, we create a surrogate system by shuffling the spike times of the causal variable. Therefore, the shuffled point process maintains the spike counts in each realization, but any consistent structure in the joint probability distribution between the two-time series is lost. Performing such random shuffling with 100 different permutations results in a distribution of DI corresponding to the null hypothesis of non-causality. The null hypothesis is rejected if the DI values estimated on the original time series are greater than the baseline threshold at a significance level of $\alpha_s = 0.05$.

To tease apart indirect influences from direct influences, we condition the DI measure on 'side' processes whose knowledge may render the involved processes statistically independent. Two types of indirect influence topologies considered in this study are proxy and cascading influences. In a proxy influence pattern between three processes, process Y_1 influences Y_2 that in turn drives the process Y_3. In such a case, the DI measure between Y_1 and Y_3 is likely to identify a causal connection. However, the two processes can be rendered statistically independent given causal knowledge of $Y_2 \rightarrow Y_3$. Similarly, a cascading topology is where two processes Y_1 and Y_3 are commonly driven by a third process, Y_2. The indirect connectivity between the two could be accounted for by their dependence on Y_2.

Causal conditioning of DI on 'side processes' is performed based on the fact that X directly, causally influences Y, iff $I(X^n \rightarrow Y^n || Z^n) > 0$. The problem of significance testing is also critical to this analysis as the null hypothesis distribution of the conditional DI measure is variable, and their estimates between independent processes do not equate to zero. To identify direct connectivities in the network, we evaluate the decrease in estimated influence between two nodes on introducing a side variable. Firstly, we identify potential indirect influences as a set of three nodes involved in cascading or proxy-type dependencies in the estimated directed graph. Across all combinations of existing dependencies between the three nodes, we evaluate the percentage change in the DI measure upon conditioning on the third node. On two edges that terminate on the same node, if an edge registers a significantly larger change in the DI measure is labeled as indirect and removed from the network. Our current implementation of causal graph construction can be extended to include verifications on more dependencies. Such an extension will merit future work. Finally, a MATLAB implementation of our DI model can be found at `https://github.com/shailajaAkella/Directed-Information`.