DIRECTED NETWORK INFERENCE USING A MEASURE OF DIRECTED INFORMATION

Ying Liu, Selin Aviyente

Michigan State University
Department of Electrical and Engineering
East Lansing, MI 48824, USA

ABSTRACT

The concept of mutual information (MI) has been widely used for inferring complex networks such as genetic regulatory networks. However, the MI based methods cannot infer directed or dynamic networks. In this paper, we propose a new network inference algorithm to infer directed acyclic networks which can determine both the connectivity and causality between different nodes based on the concept of directed information (DI) and conditional directed information. The proposed method is applied to both simulated data and Electroencephalography (EEG) data to evaluate its effectiveness.

Index Terms— Information theory, Directed graphs, Electroencephalography

1. INTRODUCTION

In many complex systems, it is important to quantify the causal relationships between its subsystems and model the interactions such as in genomic networks [1]. Information-theoretic approaches have been widely used for the inference of large networks in the bioinformatics community. Most of these methods rely on estimating the mutual information between variables from expression data to quantify the dependency. Chou-Liu tree algorithm was the first to adopt the mutual information in probabilistic model design with minimum spanning tree, which has a low number of edges even for non-sparse target networks [2]. The Relevance network (RELNET) extended these ideas by determining the relevant connections such that a pair of genes $X$ and $Y$ are connected when the mutual information is above a threshold. However, this method may infer false connections when two nodes are indirectly connected through an intermediate node [3]. The algorithm for the reconstruction of accurate cellular networks (ARACNE) addresses this problem by using the data processing inequality for mutual information [4]. All of these methods focus on quantifying the dependency between variables without taking temporal information into account. Therefore, they are limited to inferring undirected or stationary networks. In this paper, we propose a directed acyclic network inference algorithm based on estimating directed information between signals over time in order to quantify both the connectivity and causality in networks.

2. DIRECTED INFORMATION AND CAUSALLY CONDITIONAL DIRECTED INFORMATION

2.1. Directed information

The directed information measure introduced by Massey [5] is defined for two length N sequences $X^N = X_1, \cdots, X_N$ and $Y^N = Y_1, \cdots, Y_N$ as follows:

$$DI(X^N \rightarrow Y^N) = \sum_{n=1}^{N} I(X^n; Y_n|Y^{n-1}), \quad (1)$$

where $I(X; Y)$ is the mutual information between two random variables $X$ and $Y$.

Since $0 < DI(X^N \rightarrow Y^N) < I(X^N; Y^N) < \infty$, in practice a normalized version of DI, which maps DI to the [0, 1] range is used for comparing different interactions [1]:

$$
\rho_{DI} = \sqrt{1 - e^{-2\sum_{i=1}^{N} I(X^n; Y_n|Y^{n-1})}}. \quad (2)
$$

2.2. Causally conditional directed information

Based on the definition of directed information, if $X$ is the driver of $Y$, then $DI(X \rightarrow Y) > DI(Y \rightarrow X)$ which indicates causality between the two processes. However, the opposite does not hold, i.e., given a large DI value, $X$ and $Y$ can be either directly connected or indirectly through other nodes. For this reason, we introduce the concept of conditional directed information to discriminate between direct connections and the indirect ones.

The directed information from $X^N$ to $Y^N$ when causally conditioned on the sequence $Z^N$ is defined as [6],

$$DI(X^N \rightarrow Y^N | Z^N) = \sum_{n=1}^{N} I(X^n; Y_n|Y^{n-1}Z^n), \quad (3)$$

where $I(X; Y|Z)$ is the conditional mutual information between two random variables $X$ and $Y$ given $Z$.

Any three nodes in a network can interact with each other through two possible scenarios: three nodes are positioned as a chain $X \rightarrow Z \rightarrow Y$ or interact through a hub gene $X \leftarrow Z \rightarrow Y$. All three pairs $(X, Y), (Y, Z), (X, Z)$ have large DI values. However, in both cases, given the information of the intermediate or hub node $Z$ for the whole time series, $X$ and $Y$ become independent for all time points, i.e. the conditional directed information $DI(X^N \rightarrow Y^N | Z^N)$ equals to zero as will be shown by the following theorem.
Definition 1. \(X\) and \(Y\) are conditionally independent given \(Z\) if \(P(X, Y | Z) = P(X | Z)P(Y | Z)\), where \(P(X | Z)\) is the conditional probability density function of \(X\) given \(Z\).

Theorem 1. If \(X_m\) and \(Y_k\) are conditionally independent given \(Z^n\) for \(m, k = 1, \ldots, n\), and \(n = 1, \ldots, N\), then \(DI(X^N \rightarrow Y^N \parallel Z^N) = 0\).

In order to prove this theorem, we will use the following two lemmas.

Lemma 1. If \(P(X_m, Y_k | Z^n) = P(X_m | Z^n)P(Y_k | Z^n)\) for \(m, k = 1, \ldots, n\), then \(I(X_m; Y_k | Z^n) = 0\).

The proof of this lemma is given in [7].

Lemma 2. \(I(X^m; Y^k | Z^n) = 0\) iff \(I(X_m; Y_k | Z^n) = 0\) for \(m, k = 1, \ldots, n\).

Proof. First, we will prove that when \(I(X^m; Y^k | Z^n) = 0\), \(I(X_m; Y_k | Z^n) = 0\).

\[
I(X^m; Y^k | Z^n) = E_{P(X^m, Y^k, Z^n)} \log \frac{P(X^m, Y^k | Z^n)}{P(X^m | Z^n)P(Y^k | Z^n)}.
\]

When the conditional mutual information defined above is equal to 0, \(P(X^mY^k | Z^n) = P(X^m | Z^n)P(Y^k | Z^n)\). Therefore, the marginal probability can be written as,

\[
P(X_i; Y_j | Z^n) = \int P(X^mY^k | Z^n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_m
\]

\[
\times dy_1 \cdots dy_{j-1} dy_{j+1} \cdots dy_k
\]

\[
= \int P(X^m | Z^n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_m
\]

\[
\times \int P(Y^k | Z^n) dy_1 \cdots dy_{j-1} dy_{j+1} \cdots dy_k
\]

\[
= P(X_i | Z^n)P(Y_j | Z^n)
\]

Therefore,

\[
I(X_m; Y_k | Z^n) = E_{P(X^m, Y^k, Z^n)} \log \frac{P(X_m, Y_k | Z^n)}{P(X_m | Z^n)P(Y_k | Z^n)} = 0
\]

Similarly, we can prove that the converse is true.

Using Lemmas 1 and 2, we will give the proof of Theorem 1.

Proof. If \(X_m\) and \(Y_k\) are conditionally independent given \(Z^n\) for \(m, k = 1, \ldots, n\), then \(P(X_m, Y_k | Z^n) = P(X_m | Z^n)P(Y_k | Z^n)\). From Lemma 1, \(I(X_m, Y_k | Z^n) = 0\), and from Lemma 2, \(I(X^m; Y^k | Z^n) = 0\). Therefore, if \(m = n, k = n\), then \(I(X^n; Y^n | Z^n) = 0\), which is equal to \(H(X^n | Y^n | Z^n) = H(X^n | Y^n | Z^n)\) from the definition of mutual information; Similarly, if \(m = n, k = n - 1\), then \(I(X^n; Y^{n-1} | Z^n) = 0\), which is equal to \(H(X^n | Y^{n-1} | Z^n) = H(X^n | Y^n | Z^n)\). Therefore,

\[
DI(X^N \rightarrow Y^N \parallel Z^N) = \sum_{n=1}^{N} I(X^n; Y_n | Y^{n-1}_n Z^n)
\]

\[
= \sum_{n=1}^{N} [H(X^n | Y^{n-1} Z^n) - H(X^n | Y^n Z^n)]
\]

\[
= 0
\]

3. DIRECTED INFORMATION ESTIMATION

From equations (1) and (3), the computation of DI requires the estimation of joint probabilities of high dimensional random variables over time. In this paper, directed information estimation based on multi-information [8] is used to estimate the DI and causally conditional DI. The DI and conditional DI can be written in terms of multi-information as,

\[
DI(X^N \rightarrow Y^N \parallel Z^N)
\]

\[
= \sum_{n=1}^{N} [I(X^n, Y^n) - I(X^n, Y^{n-1}) - I(X^n, Y^{n-1}, Z^n)]
\]

\[
- \sum_{n=1}^{N} [I(Y^n Z^n) - I(Y^{n-1} Z^n)],
\]

where \(I(X^n, Y^n)\) is the multi-information between 2n random variables \(X_1, \ldots, X_n, Y_1, \ldots, Y_n\) and is defined in terms of mutual information as \(I(X^n, Y^n) = I(X^n; Y^n) + I(X^n) + I(Y^n)\). The multi-information can be estimated directly from data by using adaptive partitioning method [9].

4. INFERENCE ALGORITHM

4.1. Algorithm

The proposed inference algorithm utilizes both directed information and conditional directed information to infer the direct interactions between nodes. In the first step (lines 3-12), we calculate the time averaged directed information between two nodes \(i\) and \(j\). If \(DI_{ij}\) is larger than a threshold \(t_0\), the information flow from node \(i\) to \(j\) is assumed to be significant. The threshold, \(t_0\), is determined using surrogate data sets generated by Fourier bootstrapping method such that the 95% of the DI values from the random network are lower than \(t_0\). In the second step of the algorithm (lines 13-26), we determine the directionality of the connections by applying the sign test to the difference in DI between the two directions. If the hypothesis that there is no difference between the two directions can not be rejected at the 5% significance level, we keep both connections. Otherwise, we keep the connection with the larger DI. In order to identify the direct vs. indirect connections, for each node pair \(i, j\), we evaluate the conditional directed information given any other node \(k\) for all \(k \neq i, j\). If node \(k\) does not interact with nodes \(i\) and \(j\), then \(DI(X_i \rightarrow X_j | X_k)\) should be close to \(DI(X_i \rightarrow X_j)\). Otherwise, if \(k\) is an intermediate or hub node between \(i\) and \(j\), \(DI(X_i \rightarrow X_j | X_k)\) should be close to 0. Finally (lines 27-37), we compute \(DI(X_i \rightarrow X_j | X_k) / DI(X_i \rightarrow X_j)\), which is close to 1 when there are no interactions between \(k\) and \(i\), \(j\), and rank order this quantity from high to low for all \(k\). The connections with the highest ranks are kept until a desired number of connections, \(ec\), is achieved.

4.2. Validation

In order to evaluate the network inference algorithm, F-score is adopted [2]. The F-score can be interpreted as a weighted
average of the precision and recall and is defined as,
\[ F = \frac{2pr}{p + r}, \quad p = \frac{TP}{TP + FP}, \quad r = \frac{TP}{TP + FP}. \] (8)

A positive label predicted by the algorithm is considered as true positive (TP) or false positive (FP) depending on whether there exists a corresponding edge in the true network or not. Similarly, a negative label can be a true negative (TN) or false negative (FN) depending on the true network. In this paper, we will compute the F-score as a function of the expected number of connections ec.

Algorithm 1 Algorithm: Directed network inference

1: Input time series for M nodes, \( ec \) is the expected number of connections;
2: Initialize \( D \in R^{n \times n}, C \in (0, 1)^{(n \times n)}, cD \in R^{n \times n} \) as zero matrices;
3: for \( i = 1 \) to \( M \) do
4:   for \( j = 1 \) to \( M \) do
5:     \( D_{i,j} \leftarrow DI(i \rightarrow j); \)
6:     if \( D_{i,j} > t_0 \) then
7:       \( C_{i,j} = 1; \)
8:     else
9:       \( C_{i,j} = 0; \)
10:   end if
11: end for
12: end for
13: for \( i = 1 \) to \( M - 1 \) do
14:   for \( j = i + 1 \) to \( M \) do
15:     if \( C_{i,j} == 1 \) and \( C_{j,i} == 1 \) then
16:       \( h = \text{sign test}(DI(i \rightarrow j), DI(j \rightarrow i)); \)
17:     if \( h == 1 \) then
18:       if \( D_{i,j} < D_{j,i} \) then
19:         \( C_{i,j} == 0; \)
20:     else
21:       \( C_{j,i} == 0; \)
22:     end if
23: end if
24: end if
25: end for
26: end for
27: for \( i = 1 \) to \( M \) do
28:   for \( j = 1 \) to \( M \) do
29:     if \( C_{i,j} == 1 \) then
30:       \( cD_{i,j} \leftarrow \min(DI(x_i \rightarrow x_j | x_k)), k \neq i,j; \)
31:     else
32:       \( cD_{i,j} \leftarrow 0; \)
33:     end if
34: end for
35: end for
36: \( c_a = cD(:,); \)
37: \( c_b = \text{sort}(c_a) \) in descending order and keep the top \( ec \) connections

5. RESULTS

5.1. Simulated data
We first test our algorithm on a synthetic network of 14 nodes. The network contains both linear and nonlinear dependencies as shown in Fig. 1(a). We generate 128 realizations of the 14 different time series [10] for each node and compute the DI over 10 time samples. The distribution of the 182 pair-wise DI values is shown in Fig. 1(b). To determine the significance threshold, we generate 100 randomized networks and the pair-wise DI is recalculated for each permutation. The distribution of DI for random networks is also shown in Fig. 1(b). We choose \( t_0 = 0.80 \) such that 95% of the DI values for the random network lies below this threshold. All of the connections above this threshold are kept. We then eliminate the bidirectional connections based on the sign test rule. Indirect connections due to intermediate and hub nodes are also removed using conditional DI. Finally, F-score is used to quantify the performance of the algorithm in terms of the number of expected connections ec as seen in Fig. 1(c). The reconstructed network with 11 expected connections using only DI and both DI and conditional DI are shown in Figs. 1(d) and 1(e), respectively. For 11 connections, the F-score reaches 0.80. Our method effectively removes most of the indirect connections.

Fig. 1. (a) The synthetic network; (b) The distribution of directed information; (c) The F-score of simulated network; (d) Reconstructed network without considering conditional DI; (e) Reconstructed network with conditional DI.
such as $12 \leftrightarrow 13$, etc., and the missed connections are due to the weak nonlinear interactions such as $8 \rightarrow 6$, etc.

5.2. EEG data
We examined EEG data from a study containing the error-related negativity (ERN). The ERN is a brain potential response that occurs following performance errors in a speeded reaction time task. Previous work indicates that there is increased information flow associated with ERN for the theta frequency band (4-7Hz) and ERN time window 25-75ms for Error responses (ERN) compared to Correct responses (CRN) [11]. We analyze data from 6 subjects for both the CRN and ERN. The DI measure is computed over a window corresponding to the ERN response, for all trials (70 trials) between 63 electrode pairs in the theta band (4-8Hz) and construct a network with 300 connections for each subject and each response type. Figs. 2(a) and 2(b) show the connections that at least 4 subjects in each group had in common. We can observe that most connections are in the central and parietal lobes for both groups as expected since this is a speeded response task which results in increased information flow in the motor cortex. However, the CRN group has more consistency, with more common connections among subjects, compared to ERN. The ERN group shows stronger information flow between central and parietal lobes, such as $C_2 \leftrightarrow CP_1$, $C_1 \leftrightarrow CP_2$, $C_2 \leftrightarrow CP_z$, $CP_z \leftrightarrow P_1$ in accordance with prior results [11]. The bidirectional connections are mostly due to volume conduction in the brain and the short-time window used for analysis which makes it hard to determine the directionality of the information flow.

6. CONCLUSIONS
In this paper, we propose a network inference algorithm to infer directed networks using DI and conditional DI. The algorithm is applied to both synthetic data and EEG data collected and is shown to discriminate between the two response types in terms of the information flow patterns.

7. REFERENCES


