Information Theoretic Approaches for Detecting Causality in Gene Regulatory Networks

Kangkan Medhi  
Dept of Computer Sc. & Engg.  
Indian School of Mines  
Dhanbad, India  
kangkanit@gmail.com

Syed Sazzad Ahmed  
Dept of Information Technology  
North-Eastern Hill University  
Shillong, India  
sazzad.tezu@gmail.com

Swarup Roy  
Dept of Information Technology  
North-Eastern Hill University  
Shillong, India  
swarup@nehu.ac.in

Dhruba K Bhattacharyya  
Dept of Computer Sc. & Engg.  
Tezpur University  
Tezpur, India  
dkb@tezu.ernet.in

Jugal K Kalita  
Dept of Computer Sc.  
University of Colorado  
Colorado Spring, USA  
jkalita@uccs.edu

ABSTRACT

Causality detection in gene regulatory networks (GRN) is a challenging and important task. Very few techniques have been proposed so far to infer causality in GRN. A majority of them adapts information theory as a measure to infer a causal relationship. In this work we evaluate the performance of information theoretic causality detection techniques in GRN.

We consider two such measures, namely, Transfer Entropy and Interaction Information and compare their performance with Granger causality, a statistical causality inference method. For evaluation, we use synthetic gold standard data and underlying causal networks from DREAM challenges.

Experimental results reveal that Interaction Information performs better in comparison to other candidate methods for inferring causality in GRN. It is also evident from the results that performance of information theoretic approaches is sensitive towards discretization method used.

CCS Concepts

- Applied computing → Computational biology; Recognition of genes and regulatory elements; Biological networks;

Keywords

Gene Regulatory Network; Causality; Inference; Mutual Information; Transfer Entropy

1. INTRODUCTION

The genome encodes thousands of genes whose products enable numerous cellular functions. The amounts and the temporal patterns in which these products appear in the cell are crucial to the processes of life. Gene regulatory networks govern the levels of these gene products and thus play a very important role in every process of life, including cell differentiation, metabolism, the cell cycle and signal transduction.

In Gene Regulatory Network (GRN) a group of genes in a cell intercommunicate with each other and with other substances like proteins, metabolites, etc., in the cell thereby regulating the rates at which genes in the network are transcribed into mRNA [7].

Computationally, quite often GRN is represented as a graph to describe the interactions among biomolecules. A node in a graph represents a biomolecule such as a gene, a protein or a metabolite, and an edge (or link) indicates the interaction between these two biomolecules. Such interactions may be physical interactions, metabolite flow, regulatory relationships, co-expression relationships, etc. [11]. Depending upon the nature of the edges in GRN, we get two types of GRN: Directed with causal relationships and Undirected (also called as Gene Co-Expression Network). In contrast to gene co-expression networks, a GRN is a directed representation providing additional information giving direction of influence between two genes. In a GRN, causal information is an important component in deducing the existing regulatory relationship between genes or gene products. Causation describes the relationship that is present between a cause and its effect, where the later is an outcome of the former. Causation plays an important role in GRN and is depicted by a directed graph where the directed edges correspond to causal influences between gene-activities (nodes).

A number of techniques have been proposed to infer causality in general for various application domains. They are either statistical or information theoretic approaches. These methods have been applied in finding causality in GRN. Majority of them are mainly based on information theory.

In this work, we consider transfer entropy and interaction information measures for inferring causality and compare their performance with Granger causality measure.
2. MEASURES FOR INFERRING CAUSALITY IN GRN

In this section, we discuss Transfer Entropy and Interaction Information as information theoretic measures for causality detection. We also discuss a statistical causality inference method, Granger causality, which we use to compare performance with information theoretic approaches.

2.1 Interaction Information

Mutual information (MI) in a particular environment is described as the amount of information that a event contains about the occurrence of the other event [6]. However, the MI measure is symmetric and unable to infer causality. Olsen et al. in [5] used conditional mutual information with MI to infer causality in GRN. The conditional MI between random variables \( A \) and \( B \) given \( C \) is defined as:

\[
I(A, B|C) = H(A) + H(B) - H(A, B|C)
\]

where, \( H(A), H(B), H(A, B) \) are the measure of uncertainty in the values of the genes \( A, B \) and \( A \) and \( B \) taken together and conditional entropy \( H(A|B) \) quantifies the amount of information needed to describe the outcome of a random variable \( A \) given that the value of another random variable \( B \) is known.

Based on MI and conditional MI, the interaction information between the variables \( A, B \) and \( C \) is calculated as follows:

\[
T(A, B, C) = I(A, B)I(A, B|C)
\]

where, \( I(A, B) \) is the MI between \( A \) and \( B \) and \( I(A, B|C) \) is conditional MI between \( A, B \) assuming \( C \) has already occurred.

Olsen et al. in their work use the notion of \( V \)-structure. If three nodes in a network \( A, B, C \) are connected as \( A \rightarrow B \leftarrow C \), then \( B \) is called a collider and the structure is called a \( v \)-structure. A negative value of the interaction information \( T(A, B, C) \), implies that \( B \) is a collider. They use a undirected network as input and calculated interaction information to infer causality in the GRN.

2.2 Transfer Entropy

Schreiber [8] developed the notion of Transfer Entropy (TE) between two processes which measures the uncertainty reduction in determining the future state of a process by learning the past and current states of other processes.

For two stochastic processes \( X_t \) and \( Y_t \), the uncertainty reduction of \( X_{t+1} \) due to the information about the past \( \tau_Y \) states of \( Y \) is represented by

\[
Y_t^{(\tau_Y)} = Y_t, Y_{t-1}, \cdots, Y_{t-\tau_Y+1}
\]

where, \( Y_t^{(\tau_Y)} \) is the set of past \( \tau_Y \) states of \( Y \).

In addition to the information about the past \( \tau_X \) states of \( X \), represented by

\[
X_t^{(\tau_X)} = X_t, X_{t-1}, \cdots, X_{t-\tau_X+1}
\]

measured by the transfer entropy from \( Y \) to \( X \), as in [8]

\[
T_{Y \rightarrow X} = H(X_{t+1}|X_t^{(\tau_X)}) - H(X_{t+1}|Y_t^{(\tau_Y)})
\]

We can similarly define \( T_{X \rightarrow Y} \), which does not necessarily equal to \( T_{Y \rightarrow X} \). Note that \( T_{Y \rightarrow X} \) can also be interpreted as the mutual information between \( X_{t+1} \) and \( Y_t^{(\tau_Y)} \) conditioned on \( X_t^{(\tau_X)} \).

By finding out the TE between all pair of genes one can obtain an inferred causal network of genes and then apply a heuristic rule to differentiate indirect causal relations from direct ones.

Next, we discuss a non information theoretic approach as well.

2.3 Granger Causality Test

Granger Causality (GC) a statistical notion of causality based on prediction states that "If a signal \( X \) Granger-causes (or G-cause) a signal \( Y \), then past values of \( X \) should contain information that helps predict future value of \( Y \) above and beyond the information contained in past values of \( Y \) alone." [2] This means that \( X \) can help in reducing the errors which were inevitable if the calculation was done using the values of \( Y \) only.

GC is generally tested in the context of linear regression models. For illustration, a bivariate linear autoregressive model of two variables \( X_1 \) and \( X_2 \) is considered as shown below:

\[
X_1(t) = \sum_{j=1}^{n} A_{11,j} X_1(t-j) + \sum_{j=1}^{p} A_{12,j} X_2(t-j) + E_1(t)
\]

\[
X_2(t) = \sum_{j=1}^{n} A_{21,j} X_1(t-j) + \sum_{j=1}^{p} A_{22,j} X_2(t-j) + E_2(t)
\]

where \( E_1 \) and \( E_2 \) are the errors in prediction for the time series \( X_1 \) and \( X_2 \) respectively, \( n \) is the model order i.e. the maximum number of lagged observations included in the model, and \( A \) is the matrix containing the coefficients of the model (i.e., the contributions of each lagged observation to the predicted values of \( X_1(t) \) and \( X_2(t) \)).

If the variance of \( E_1 \) (or \( E_2 \)) is decreased by including the \( X_2 \) (or \( X_1 \)) terms in the first (or second) equation, then it can be concluded that \( X_2 \) (or \( X_1 \)) Granger causes or G-causes \( X_1 \) (or \( X_2 \)). In other words, \( X_2 \) G-causes \( X_1 \) if the coefficients in \( A_{12} \) are jointly significantly different from zero. This can be checked by performing an F-test of the null hypothesis that \( A_{12} = 0 \), given assumptions of covariance stationarity on \( X_1 \) and \( X_2 \). By calculating the logarithm of the corresponding F-test the magnitude of a G-causality interaction can be computed.

Next, we evaluate performance of above methods on DREAM challenge data.

3. EXPERIMENTAL EVALUATION

In this section the results of various experiments performed on time series data is presented. Then we evaluate their performances using different validation measures.

3.1 Dataset

In-silico GRNs are used for evaluating the performance of different causality detection techniques. Time series data for experimentation are obtained from DREAM (Dialogue for Reverse Engineering Assessments and Methods) network inference challenge, provided by Marbach’s Gene Net Weaver [3] platform. The challenges that are available are Dream3 and Dream4. Dream3 involves fifteen benchmark datasets, five each of sizes 10, 50 and 100. The structures of the benchmark networks are obtained by extracting modules from real biological networks. At each size, two of the networks are extracted from the regulatory network of E. coli and Yeast.
3.2 Discretization methods used

Discretization is commonly used to convert continuous data to discrete ones. Some of the measures for inferring causality in GRN like Interaction Information, Transfer entropy etc., requires discrete data as input and hence the available continuous data needs to be converted to discrete time series data. We use following techniques for pre-processing continuous time series data.

3.2.1 Equal Width (EW)

In EWFD the continuous time series data is partitioned into $k$ equally size bins where $k$ is a parameter supplied by the user. If a variable $x$ is observed to have values bounded by $x_{\text{min}}$ and $x_{\text{max}}$ then this method computes the bin width

$$\delta = \frac{x_{\text{max}} - x_{\text{min}}}{k}$$

and constructs bin boundaries, or thresholds, at $x_{\text{min}} + i\delta$ where $i = 1, \ldots, k - 1$. The method is applied to each continuous feature independently. It makes no use of instance class information whatsoever and is thus an unsupervised discretization method[1].

3.2.2 Equal Frequency (EF)

In EFD the continuous time series data is partitioned into $|\chi_i|$ intervals, each having the same number, $m/|\chi_i|$, of data points. As a result, the intervals can have different sizes. If the $|\chi_i|$ intervals have equal frequency, then the computation of entropy is straightforward: $\log \frac{1}{|\chi_i|}$. However, if one of the bins is more dense than the others, then the resulting entropy needs to be estimated. This discretization method is mentioned in [10] as one of the most efficient methods. The difference between equal width and equal frequency discretization methods are summarized here,

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Dataset</th>
<th>In silico network</th>
<th>Size of the network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dream3</td>
<td>1</td>
<td>Ecoli1</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Ecoli2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Yeast1</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Ecoli1</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Yeast1</td>
<td>100</td>
</tr>
<tr>
<td>Dream4</td>
<td>6</td>
<td>insilico1</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>insilico5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>insilico1</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>insilico5</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2: Properties of discretization methods

<table>
<thead>
<tr>
<th>Evaluation Criteria</th>
<th>Equal Width</th>
<th>Equal Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised/Unsupervised</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic/Static</td>
<td>Static</td>
<td>Static</td>
</tr>
<tr>
<td>Splitting/Merging</td>
<td>Split</td>
<td>Split</td>
</tr>
<tr>
<td>Direct/Incremental</td>
<td>Direct</td>
<td>Direct</td>
</tr>
</tbody>
</table>

3.2.3 Global Equal Width (GEW)

Global Equal Width is similar to Equal Width but in the former the considered range between the minimum and the maximum value is different from the later. In Equal Width we take only the local maximum and local minimum as the range whereas in Global Equal Width we take maximum and the minimum of the entire dataset [4].

3.3 Implementation Setup

We use R to implement all the programs. To calculate the effectiveness of the inferred networks, we compare with the actual networks generated from in-silico DREAM challenge data. Three different metrics for evaluating accuracy are used - accuracy, $F_\beta$ and AUROC (Area under Receiver-Operator Characteristics curve) score. The ROC is a comparison of two operating characteristics -True Positive Rate and False Positive Rate and hence it is also known as relative operating characteristic curve [9]. ROC curves may not be the appropriate measure when a dataset contains large skews in class distribution, which is usually the case in transcriptional network inference. We have used $\beta = 0.5$ for $F$ score.

3.4 Experimental Results

We measure the performance variation of information theoretic approaches with respect to different discretization methods. We use DREAM 4 challenge data (in-silico1-10) for such experimentation and calculate three different validity scores, $F_\beta$, ROC and accuracy. Figure 1 and 2 clearly reveals that information theoretic approaches are highly dependent of discretization methods and their parameters. The relative performance of both the methods with varying bin size (2 to 5) and discretization methods is given in Figure 3.

We also observe that Equal frequency discretization works better than the other two methods for bin size = 2 and 3. We also observe that for this combination of dataset and inference method, there is not much change as the number of bins for Equal Width discretization is varied. We compare
the performance of information theoretic approaches with GC and present the results of the experiments in Figure 4. Comparison of the result reveals that the Interaction Information method for causality inference performs consistently better than the other two methods for all datasets. However, it fails to infer positive edges.

4. CONCLUSION

In this work, we evaluate the performance of information theoretic approaches such as Interaction information or Transfer Entropy to infer causal relationship in GRNs from time-series gene expression data. We also compare their performance with a statistical causality inference method, Granger Causality. We use DREAM challenge data for experimentation. We also use different discretization methods as performance of information theoretic method depends on them. Experimental results reveals that Interaction Information measures are more effective in inferring causal relation in comparison to other candidate methods.

5. REFERENCES


