A methodology for detection of causal relationships between discrete time series on systems

Rute Souza de Abreu

rute.s.abreu@gmail.com

Departamento de Computação e Automação

Universidade Federal do Rio Grande do Norte

January 25, 2019

Contents

Introduction

- Motivation of the study
  - Causal relationships importance
- Proposal
  - Define a methodology to identify causal relationships between discrete time series on systems.
    - Transfer Entropy
    - Structural Learning on Bayesian Networks
- Related Work
  - Implementation of transfer entropy for detecting causal relationships between industrial alarm series [1]
  - Use of transfer entropy to detect neuronal connections [2]
  - Use of transfer entropy, Granger causality and Dynamic Bayesian Networks for reconstruction of genetic networks [3]

[1](SU et al., 2017) (HU et al., 2017), (YU; YANG, 2015)
[2](VICENTE et al., 2011)
[3](TUNG et al., 2007) (ZOU; FENG, 2009)

Theoretical Foundations

Information Theory

- Entropy
- Kullback-Leibler Divergence
- Mutual Information
- Transfer Entropy

Bayesian Networks

- Bayesian Networks - Concept
- K2-Algorithm
- Medium Description Length (MDL)

Theoretical Foundations - Entropy

Definition

The amount of information produced by an information source.

$$H = \sum_{i}^{n} p_i(i) log_2 \frac{1}{p(i)} \tag{1}$$

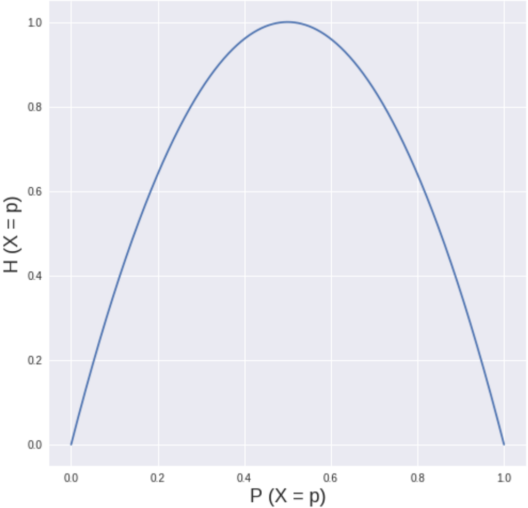
Theoretical Foundations - Entropy

It is related to the frequency of the appearance of each value or the amount of surprise obtained given the appearance of a state or value.

Properties:

- It is continuous in the domain of p_i, which is the probability mass function.
- It is monotonically increasing, in the n domain, when all events are likely equally.
- It is weighted additive when a choice is broken down into successive choices.

Theoretical Foundations - Entropy



Theoretical Foundations - Kullback-Leibler Divergence

Definition

- Given a variable I, with probability distribution p. It measures the error or divergence when it is assumed that the probability distribution of I is q, instead of p.

$$KL_I = \sum_{i}^{n} p(i) log \frac{p(i)}{q(i)} \tag{2}$$

Theoretical Foundations - Kullback-Leibler Divergence

$$K_{I,J} = \sum_{i,j}^{n} p(i,j) log \frac{p(i,j)}{q(i,j)} \tag{3}$$

$$K_{I|J} = \sum_{i,j}^{n} p(i,j) log \frac{p(i \mid j)}{q(i \mid j)} \tag{4}$$

Theoretical Foundations - Mutual Information

Definition

It is a measure derived from the Kullback-Leibler divergence. It is calculated between two processes I and J and quantifies the amount of information obtained about one process when observing the other.

It can be seen as the “information produced by erroneously assuming that the two processes are independent”.[4]

$$MI_{I,J} = \sum_{i,j}^{n} p(i,j) log \frac{p(i,j)}{p(i) \cdot p(j)} \tag{5}$$

[4](SCHREIBER, 2000)

Theoretical Foundations - Transfer Entropy

Definition

It is a theoretical measure “that shares some of the desired properties of mutual information but takes the dynamics of information transport into account”.[5]

$$TE_{J->I} = \sum_{i,i_{t+h},j} p(i_{t+h}, i_t^k, j_t^l) log \frac{p(i_{t+h} \mid i_t^k, j_t^l)}{p(i_{t+h} \mid i_t^k)} \tag{6}$$

[5] (SCHREIBER, 2000)

Theoretical Foundations - Transfer Entropy

$$i^k = [i_t, ..., i_{t-k+1}] \tag{7}$$

$$j^l = [j_t, ..., j_{t-l+1}] \tag{8}$$

- It infers the veracity of the equation:

$$p(i_{t+h} \mid i_t^k, j_t^l) = p(i_{t+h} \mid i_t^k) \tag{9}$$

Theoretical Foundations - Transfer Entropy

Understanding of the parameters (Example)

- Series size = 6 samples
- k = 3 samples
- l = 2 samples
- h = 1 sample

I series

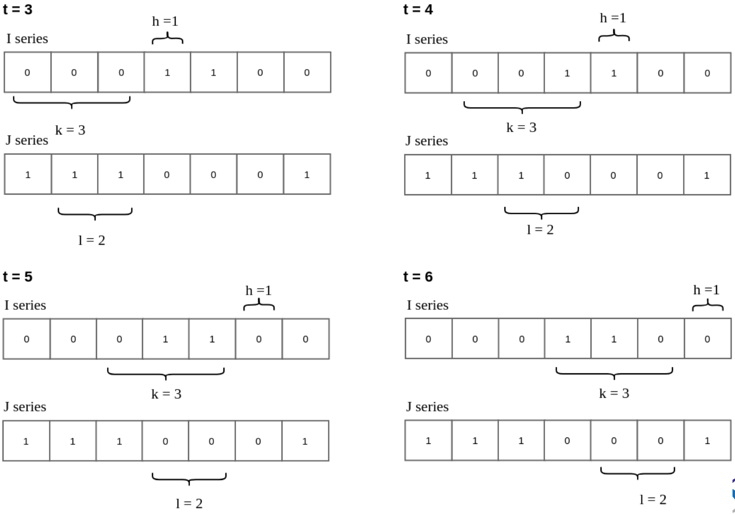
0	0	0	1	1	0	0

J series

1	1	1	0	0	0	1

Figure 2: I and J series.

Theoretical Foundations - Transfer Entropy



Theoretical Foundations - Bayesian Networks

- Probabilistic models based on direct acyclic graphs.
- The nodes represent variables of interest, while the connections represent relationships of informational or causal dependencies.[6]
- Each variable is independent from its nondescendants, given its parents.

[6](PEARL, 2011)

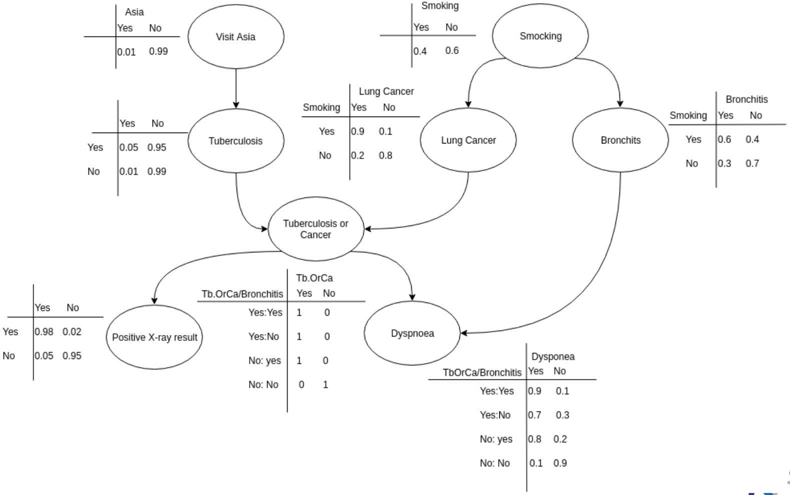
- It factors the joint the distribution of the variable into local joint distributions.

$$P(A_n, \cap \ldots \cap A_1) = \prod_{i}^{n} P(A_i \mid \cap_{j=1}^{i-1} A_j) \tag{10}$$

$$P(A_n, \cap \ldots \cap A_1) = \prod_{i=1}^{n} P(A_i | pa_i) \tag{11}$$

$$P(\,TbOrCa \mid Cancer,\ \ Bronch.,\ \ Asia) = P(\,TbOrCa \mid Cancer) \tag{12}$$

Theoretical Foundations - Bayesian Networks



Theoretical Foundations - K2 Algorithm

K2 - Algorithm

- Bayesian method for estimating a probabilistic network from data.
- The algorithm searches for the network that has the highest posterior probability given a database of records.
- Necessity of a prior ordering on nodes.
- Scores each local network in order to find the most probable structure.

K2-Algorithm

Case	Variable values for each case x_1	x_2	x_3
1	*present*	*absent*	*absent*
2	*present*	*present*	*present*
3	*absent*	*absent*	*present*
4	*present*	*present*	*present*
5	*absent*	*absent*	*absent*
6	*absent*	*present*	*present*
7	*present*	*present*	*present*
8	*absent*	*absent*	*absent*
9	*present*	*present*	*present*
10	*absent*	*absent*	*absent*

Figure 5: Database of cases

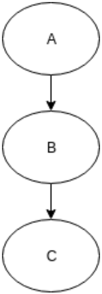


Figure 6: Possible structure.

K2-Algorithm

Algorithm 1: K2 Algorithm

Input: a set of n nodes - $(x_1 \ldots x_n)$
an ordering for the nodes ,
a data set of n columns and m cases - database

Output: Bayesian Network Topology

$\pi_i \leftarrow$ parents of node i

for $i \leftarrow 1$ **to** n **do**
  $\pi_i \leftarrow \emptyset$
end
for $i \leftarrow 1$ **to** n **do**
  $P_{old} \leftarrow g(i, \pi_i)$
  while *True* **do**
    $pred_{x_i} \leftarrow$Pred(x_i) - set of nodes that precede x_i
    select the node $x_j \in pred_{x_i} \setminus \pi_i$ *that maximizes:* g(i,$\pi_i \cup \{x_j\}$)
    $P_{new} \leftarrow$ g(i,$\pi_i \cup \{x_j\}$)
    sigma $\leftarrow P_{new} > P_{old}$
    if *sigma* = *True* **then**
      $P_{old} \leftarrow P_{new}$
      $\pi_i \leftarrow \pi_i \cup x_j$
    end
    if ***not***$(pred_{x_i} = \emptyset)$ **then**
      $pred_{x_i} \leftarrow pred_{x_i} \setminus x_j$
    end
    if *(not sigma or* $pred_{x_i} = \emptyset$*)* **then**
      break
    end
  end
  returns the parent set of each node
end

Figure 7: K2-Algorithm

Proposed Methodology

General Concept

By the use of the concepts of Information Theory and Bayesian Networks it is intended to unite the method of Transfer Entropy and the K2 algorithm in order to generate a single methodology for the detection of causal relationships.

Approach

To model the system as a graph, in which the nodes will be the entities related to each other, by a causality reelationship. This detection will be made in five stages

Proposed Methodology

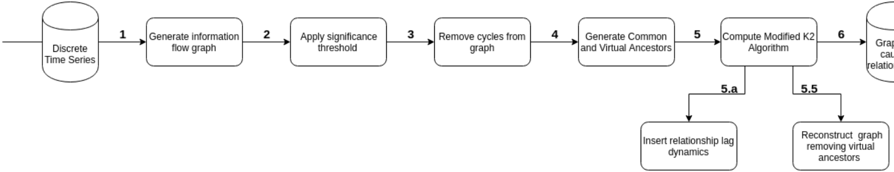


Figure 8: Proposed Methodology

Generation of graph of information flow

- Given a system with N variables, it computes the Transfer Entropy pairwise for the set of variables.
- For each pair of variables it computes the method *h* times.
- Chooses the greatest entropy value from the *h* computations.

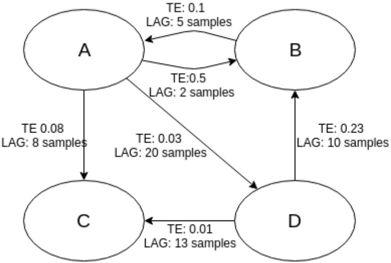
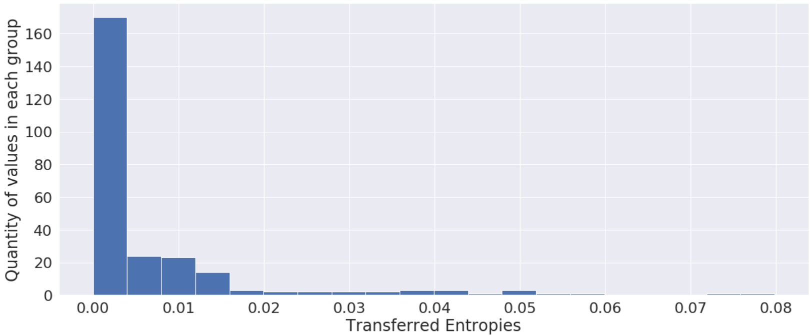


Figure 9: Example of output from stage 1

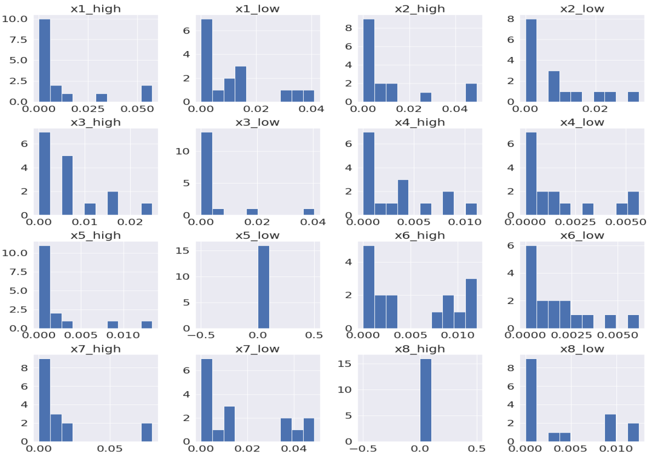
Application of Statistical Threshold

- Because the Transfer Entropy is calculated pairwise, it can produce a dense graph. Since due to the noise cointained in the time series,the chance of a zero entopy becames low.
- This stage aims, to the define a threshold for the entropy values based on the distribution of the computed values.

Application of Statistical Threshold



Application of Statistical Threshold



Removal of Cycles

- In order to prepare the graph to be used on K2, third stage does a removal of cycles.
- It consists of recursive search on the graph, in which the nodes “above” the node-cause, are marked as its ancestors. Therefore, he ancestors of a node-cause cannot be related to it as an effect.
- The removal is made based on the criterion of the higher information

Removal of cycles

Generation of Common and Virtual Ancestors

What is a common and a virtual ancestor?

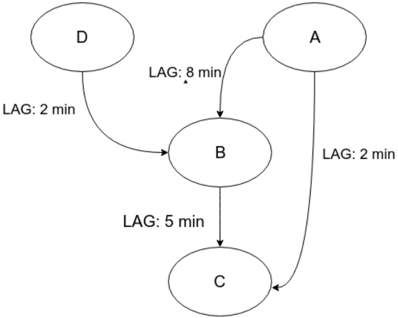


Figure 10: Ordinary Graph

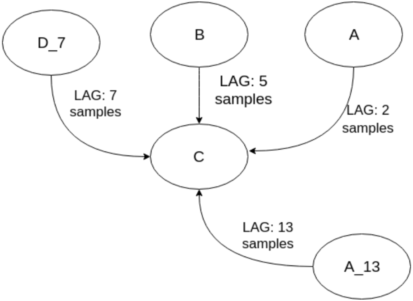


Figure 11: Modified Graph

Prior Strucuture to K2

Modification and Computation of K2

Case Study and Results

Resumo

References I

HU, Wenkai et al. Cause-effect analysis of industrial alarm variables using transfer entropies. **Control Engineering Practice**, Elsevier, v. 64, p. 205–214, 2017.

PEARL, Judea. Bayesian networks. 2011.

SCHREIBER, Thomas. Measuring information transfer. **Physical review letters**, APS, v. 85, n. 2, p. 461, 2000.

SU, Jianjun et al. Capturing Causality for Fault Diagnosis Based on Multi-Valued Alarm Series Using Transfer Entropy. **Entropy**, Multidisciplinary Digital Publishing Institute, v. 19, n. 12, p. 663, 2017.

References II

- TUNG, Thai Quang et al. Inferring gene regulatory networks from microarray time series data using transfer entropy. In: IEEE. COMPUTER-BASED Medical Systems, 2007. CBMS'07. Twentieth IEEE International Symposium on Computer-Based Medical Systems. Maribor, Slovenia: IEEE, 2007. p. 383–388.
- VICENTE, Raul et al. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. **Journal of computational neuroscience**, Springer, v. 30, n. 1, p. 45–67, 2011.
- YU, Weijun; YANG, Fan. Detection of causality between process variables based on industrial alarm data using transfer entropy. **Entropy**, Multidisciplinary Digital Publishing Institute, v. 17, n. 8, p. 5868–5887, 2015.

References III

ZOU, Cunlu; FENG, Jianfeng. Granger causality vs. dynamic Bayesian network inference: a comparative study. **BMC bioinformatics**, BioMed Central, v. 10, n. 1, p. 122, 2009.