Causality Challenge

Pot-Luck

Bring Your Own Problem

Isabelle Guyon, Clopinet
Constantin Aliferis and Alexander Statnikov, Vanderbilt Univ.
André Elisseeff and Jean-Philippe Pellet, IBM Zürich
Gregory F. Cooper, Pittsburg University
Peter Spirtes, Carnegie Mellon
Causality Workbench

• **Goal:** Benchmark causal discovery algorithms.

• **Method:**
  – Challenges.
  – Repository of datasets, tasks, models, software, etc.
  – Interactive workbench.
  – Weekly teleconference seminar.

• **So far…**
  – Causality Challenge #1: Causation and Prediction (WCCI 2008).
  – Causality Challenge #2: Pot-luck (NIPS 2008).

• **Winter, 2008:** Start developing an interactive workbench.

• **June, 2009:** KDD workshop on causality in time series?
Why a new challenge?

• **Causality challenge #1**
  – Favor “depth”
    • Single well defined task
    • Rigorous performance assessment

• **Causality challenge #2**
  – Favor “breadth”
    • Many different tasks
    • Encourage creativity
http://clopinet.com/causality

Causality Workbench

PROMO: Simple causal effects in time series
Contact: Jean-Philippe Pellet - Submitted: 2008-09-15 18:37 - Views: 63

The PROMO dataset proposes the task to identify which promotions affect sales. Artificial data about 1000 promotion variables and 100 product sales is provided. The goal is to predict a 1000x100 boolean influence matrix, indicating for each (i,j)...

Authors: Causality workbench team
Key facts: This dataset contains artificial data about product sales and promotions as time series. There are 1000 binary promotions variables and 100 continuous product...
Keywords: time.series, structural.equation.models

SIGNET: Abscisic Acid Signaling Network
Contact: Jerry Jenkins - Submitted: 2008-09-17 21:53 - Views: 20

The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynamic simulations of the true rules, using an asynchronous...

Authors: Jerry W. Jenkins, Abhishek Soni
Key facts: Simulated data with a Boolean network modeling a biological signaling network. Time series of 21
### Pot-Luck challenge

<table>
<thead>
<tr>
<th>Task</th>
<th>Participants (views)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYTO</td>
<td>2 (394)</td>
<td>real self eval</td>
</tr>
<tr>
<td>LOCANET</td>
<td>10 (558)</td>
<td>real artif</td>
</tr>
<tr>
<td>PROMO</td>
<td>3 (570)</td>
<td>artif self eval</td>
</tr>
<tr>
<td>SIGNET</td>
<td>2 (415)</td>
<td>artif</td>
</tr>
<tr>
<td>TIED</td>
<td>1 (330)</td>
<td>artif</td>
</tr>
<tr>
<td>CauseEffectPairs</td>
<td>5 (218)</td>
<td>real</td>
</tr>
<tr>
<td>Stemmatology</td>
<td>0 (109)</td>
<td>real self eval</td>
</tr>
</tbody>
</table>
**Other donated datasets**

<table>
<thead>
<tr>
<th>Task</th>
<th>Views</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>WebLogs</td>
<td>90</td>
<td>real</td>
</tr>
<tr>
<td>MIDS</td>
<td>65</td>
<td>artif</td>
</tr>
<tr>
<td>NOISE</td>
<td>43</td>
<td>real</td>
</tr>
<tr>
<td>SECOM</td>
<td>59</td>
<td>real</td>
</tr>
<tr>
<td>SEFTI</td>
<td>35</td>
<td>real</td>
</tr>
</tbody>
</table>

[http://clopinet.com/causality](http://clopinet.com/causality)
Winners

PRIZES
Best benchmark result: Kun Zhang and Aapo Hyvärinen
Best contributed task: Guido Nolte

MENTIONS
Significant advance on

REGED: Ernest Mwebaze and John Quinn
SIDO: You Zhou, Changzhang Wang, Jianxin Yin, Zhi Geng
      Mehreen Saeed
SIGNET: Cheng Zheng and Zhi Geng
SIGNET: Advanced Analytics, Intel, LTD
TIED: 
Causal discovery from real manipulations

The CYTO problem

Karen Sachs et al
What if we cannot experiment?

The LOCANET problem

• Experiments may be infeasible, costly or unethical.
• Using only observations we may want to predict the effect of new policies.
• Policies may consist in manipulating several variables.
• Task: Find the local causal structure around a given target variable (depth 3 network) in four datasets (REGED, CINA, SIDO, and MARTI).
**Multiple alternative solutions**

**The TIED problem**

Two disjoint subsets of variables $V_1$ and $V_2$ are **Target Information Equivalent w.r.t. target $Y$** $\text{TIE}_Y(V_1, V_2)$, iff:

- $V_1 \perp Y$
- $V_2 \perp Y$
- $V_1 \perp Y \mid V_2$
- $V_2 \perp Y \mid V_1$

$\begin{array}{c}
\text{TIE}_Y(X_1, X_2) \\
\text{TIE}_Y(X_1, X_3) \\
\text{TIE}_Y(X_1, X_{11}) \\
\text{TIE}_Y(X_2, X_3) \\
\text{TIE}_Y(X_2, X_{11}) \\
\text{TIE}_Y(X_3, X_{11})
\end{array}$
A dynamical system: SIGNET
Abscisic Acid Signaling Network

Donated by Jenkins and Soni
Resimulated from Li, Assmann, Albert, PLOS, 2006
Causality in time series: PROMO
A simulated marketing task

- 100 products
- 1000 promotions
- 3 years of daily data
- **Goal:** quantify the effect of promotions on sales.

The difficulties include:
- non iid samples
- seasonal effects
- promotions are binary, sales are continuous

Jean-Philippe Pellet & André Elisseeff
Causal direction among only two variables?  
*The CauseEffectPairs problem*

- Many causal discovery methods rely on tests of conditional independence between 3 or more variables.
- **Task:** Find the causal direction among pairs of variables (real data, e.g. temperature and altitude).

*Dominik Janzing*
What’s next?

• Proceedings of NIPS workshop (JMLR, early 2009):

• Depth vs. breadth → focus months:
  – Teleconference presentations on one particular challenge.
  – Deadline for submission; result analysis and debate.

• Causality challenge #3:
  – Focus on time series.
  – Target KDD, June 2009.

• Interactive workbench:
  – Under development; target next NIPS.

http://clopinet.com/causality