Causal Learning Without DAGs

David Duvenaud, Daniel Eaton, Mark Schmidt and Kevin Murphy
University of British Columbia

Abstract

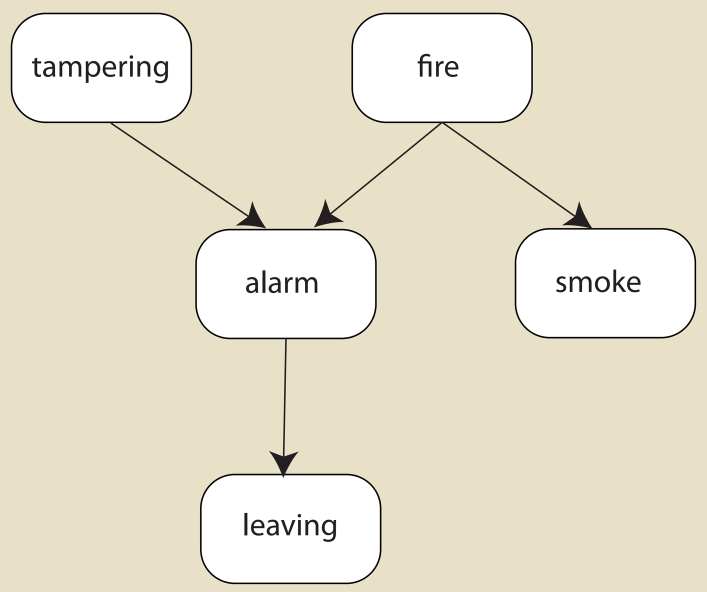
Causal learning methods are often evaluated in terms of their ability to discover a true underlying DAG structure. However, in general the true structure is unknown and may not be a DAG. We therefore consider evaluating causal learning methods in terms of predicting the effects of interventions on unseen test data. Given this task, we show that there exist a variety of approaches to modeling causality, generalizing DAG-based methods. Our experiments on synthetic and biological data indicate that some non-DAG models perform as well or better than DAG-based methods at causal prediction tasks.

Background

Goal:

To predict the results of unobserved actions on a system. Formally, to predict P(X|A) where X are the observed variables, and A is a previously unseen action.

Why Causal DAGs are popular for this task:



A causal DAG has an arrow from A to B iff A causes B. Pearl [1] showed that under some circumstances, causal DAGs are the appropriate tool to use to model interventions.

If interventions are 'perfect' (affect only one node) and affect only observable nodes, then modeling interventions can be done correctly by performing "graph surgery": Cutting the arcs leading into the intervened node, and conditioning on the intervened value as in a normal DAG.

Using causal DAGs in this way, one can predict the effects of novel interventions on a system.

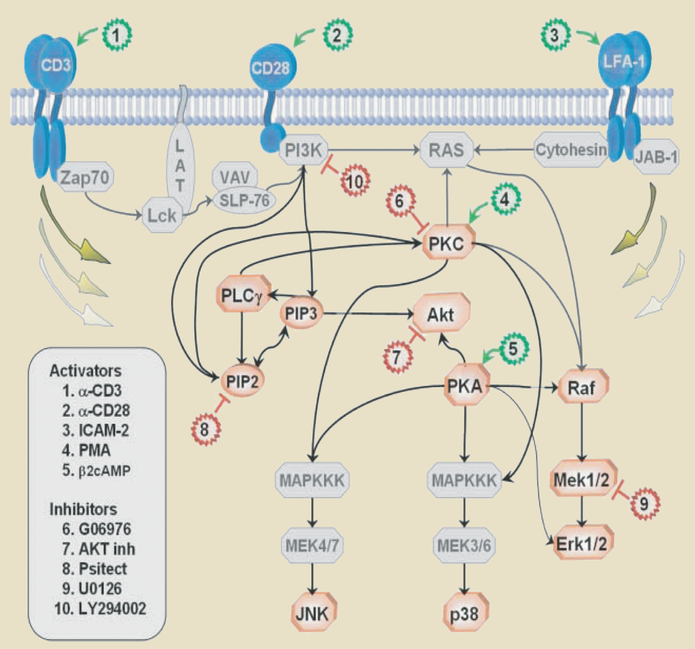
Caveats:

- In general, observational data will only recover graph structure up to Markov equivalence.
- Graph surgery operation is only applicable if interventions perfectly affect only visible nodes.
- The true independence structure may have indendencies not representable by a DAG. [4]

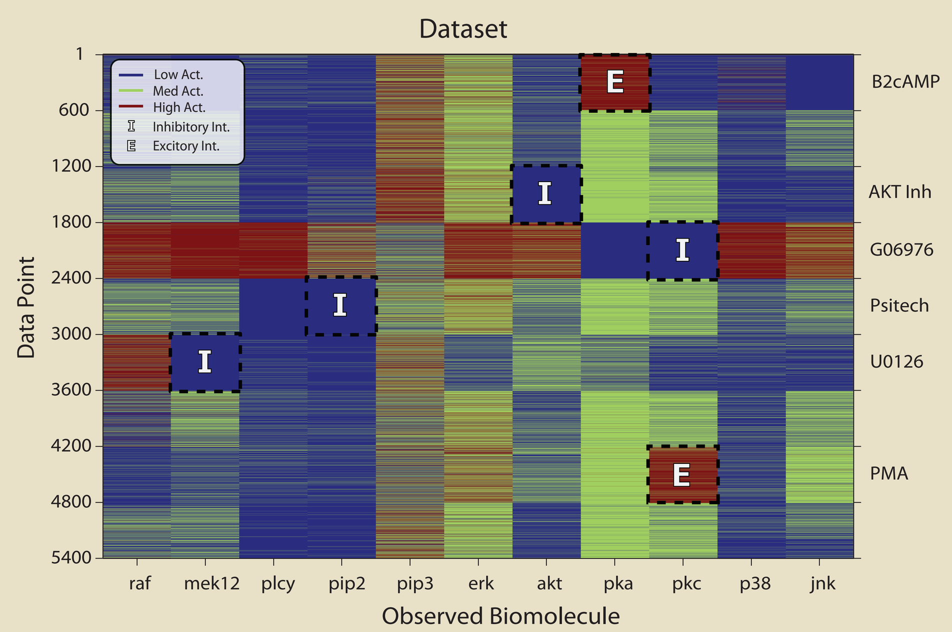
If any of these caveats hold, we may be better off using another model.

Biological Dataset

- Biological dataset of Sachs *et al.* [3] measuring protein concentration levels in a T-cell signalling pathway.
- 11 attributes discritized to 3 levels, 5400 records, 6 experimental conditions (including pure observation and 5 different interventions).
- Interventions were designed to target single proteins.



"Ground truth" (Source: [3])



Visualization of the dataset. "I" and "E" boxes represent interventions on specific proteins. (Source: [3])

Pooling Data Across Actions

We examined 4 ways to generate a conditional model of the data X given the action A:

I) Ignore Actions

We simply ignore A and build a generative model of P(X). This has the advantage that we gain statistical strength by pooling data across the actions, but has the disadvantage that we make the same prediction for all actions.

II) Fit Independent Models for Each Action

We fit a separate model P(X|A) for each unique joint configuration of A. This is advantageous over the ignore model in that it makes different predictions for different actions, but the disadvantage of this model is that it does not leverage information gained between different action combinations, and can not make a prediction for an unseen configuration of A.

III) Fit a Model Conditional on Actions

We build a model of P(X|A), where we use some parametric model relating the A's and X's. This will allow us to borrow strength across action regimes, and to handle novel actions.

IV) Assume Perfect Interventions

We assume perfect interventions, and find the MAP DAG to represent the causal structure of the system. Since we know the targets of the interventions, we can do inference for new actions by Pearl's graph-surgery method.

Alternate Models

I) Mixture Models

In the Ignore and Independent scenarios, we fit a Mixture of Independent Multinomials (A.K.A. Mixture of Bernoullis).

In the conditional case, we fit a mixture of independent logistic regressors. (Similar to a Mixture of Experts)

II) Undirected Graphical Models

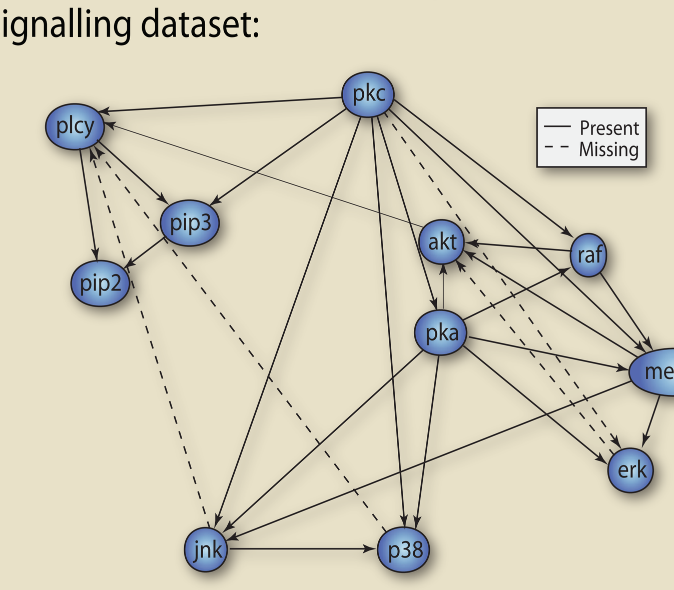
When modeling P(X), we construct a Markov Random Field, with factors for each X_i node and each $X_i - X_j$ edge.

For modeling P(X|A), we construct a Conditional Random Field, in which we additionally create factors for each $X_i - X_j$ edge, and for each (X_i, X_j, A_k) triple.

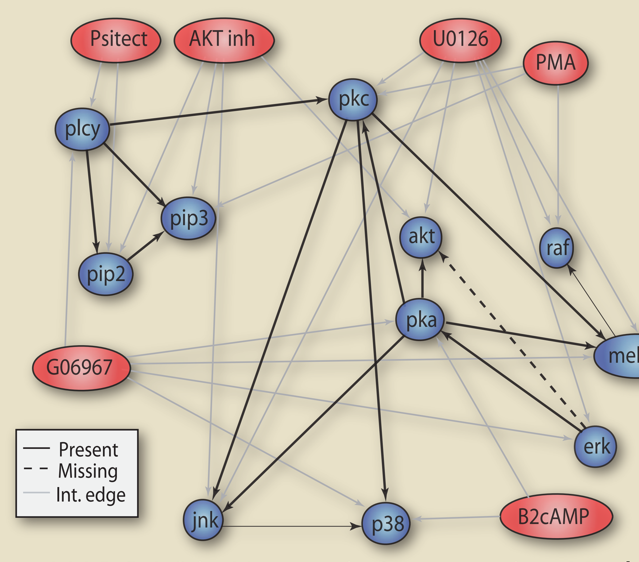
III) Uncertain Intervention DAG

Eaton et al. [2] showed that if interventions are not perfect, we can learn the targets and effects of each action, by learning an expanded graph. The graph is expanded by adding a new node for each action.

Shown here are results of the perfect intervention and uncertain intervention DAG models on the protein signalling dataset:



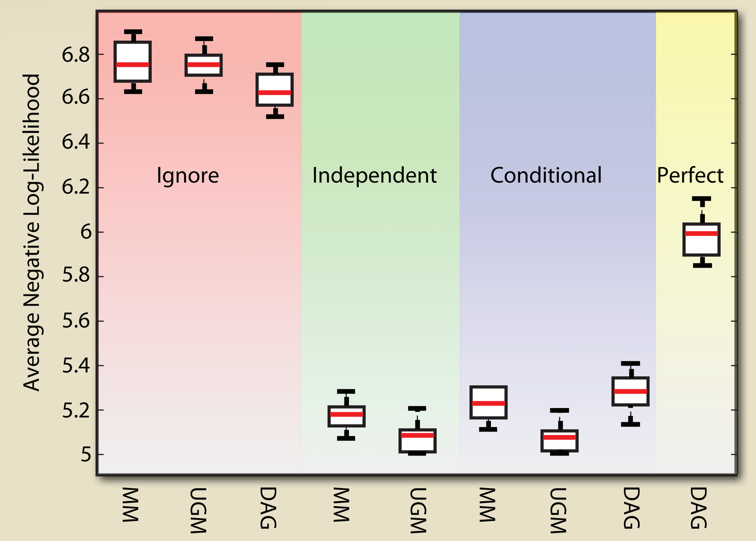
MAP DAG in Perfect Intervention Model:
Has more edges than biological model



MAP DAG in Uncertain Intervention Model:
Data suggest that interventions were not perfect, and affected multiple nodes.

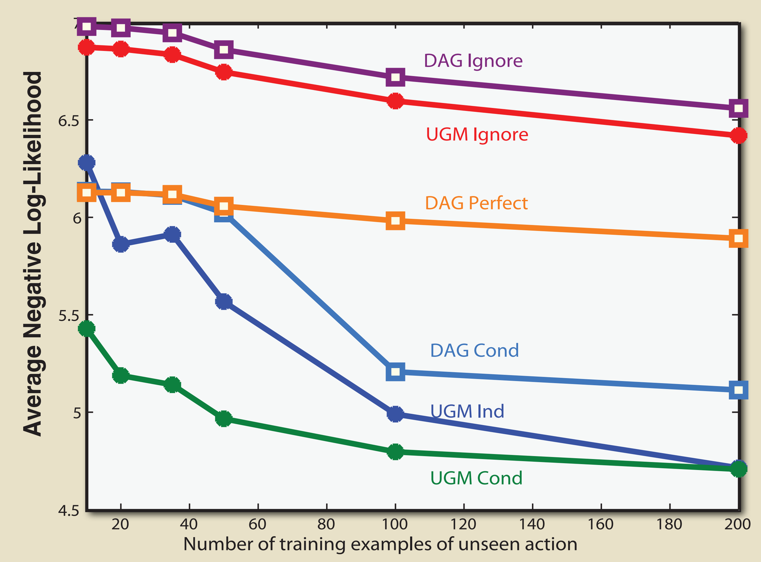
Results on Protein-signalling Dataset

Predictive Accuracy On Previously Observed Actions



- Even though the biological interventions were designed to only affect one target, the perfect intervention model performs relatively poorly. Learning the targets of actions (Conditional DAG) performs much better.
- All Independent and Conditional methods perform similarily.
- All models ignoring the actions performed poorly.

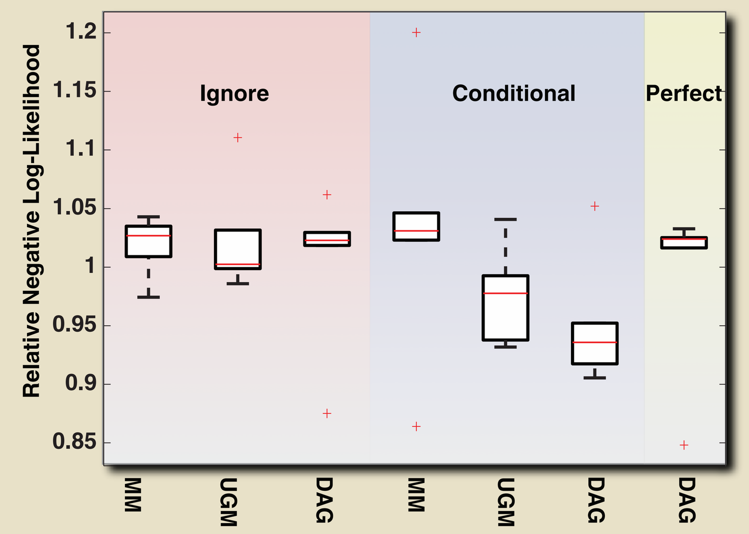
Predictive Accuracy on Rarely Seen Actions



- Conditonal UGMs need less data to learn a new action than independent UGMs

Thus, statistical strength can be borrowed across actions, even without knowing the targets of interventions.

Predictive Accuracy on Completely Unseen Actions



- All methods perform similarily, even those that ignore actions.
- The perfect DAG model, designed specifically to predict the effects of new actions, performs slightly worse than some conditional models.

Similar results hold for synthetic datasets sampled from Structural Equation Models.

Conclusions

- Causal learning can be viewed as the task of modeling the effects of unseen actions, as opposed to finding the 'true causal structure' of a system.
- Many different models besides DAGs are effective at this task.

References

[1] J. Pearl. Causality: Models, Reasoning and Inference. Cambridge Univ. Press, 2000.
[2] D. Eaton and K. Murphy. Exact Bayesian structure learning from uncertain interventions. In AI/Statistics, 2007.
[3] K. Sachs, O. Perez, D. Pe'er, D. Lauffenburger, and G. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308, 2005.
[4] A. P. Dawid. Beware of the DAG! Journal of Machine Learning Research, 2009. To appear.

To appear in the Journal of Machine Learning Research special issue on Causality