

# Learning Belief Network Structure from Data Under Causal Insufficiency

Mieczyslaw A. Klopotek

Institute of Computer Science, Polish Academy of Sciences

PL 01-237 Warsaw, 21 Ordona St., Fax: (48-22) 37-65-64

Phone: (48-22) 36-28-85 ext. 45, e-mail: klopotek@plearn.bitnet

**Abstract.** Hidden variables are well known sources of disturbance when recovering belief networks from data based only on measurable variables. Hence models assuming existence of hidden variables are under development. This paper presents a new algorithm exploiting the results of the known CI algorithm of Spirtes, Glymour and Scheines [4]. CI algorithm produces partial causal structure from data indicating for some variables common unmeasured causes. We claim that there exist belief network models which (1) have connections identical with those of CI output, (2) have edge orientations identical with CI (3) have no other latent variables than those indicated by CI, and (4) and the same time fit the data. We present a non-deterministic algorithm generating the whole family of such belief networks.

## 1 Introduction

It is a well known phenomenon of human mind to think in terms of causality. The background behind this paradigm is a strong belief that an event may in fact have only few causes so that reasoning about real world events may be kept from explosion of alternative explanations by identifying intrinsic causality. Belief networks (BN), bayesian networks, causal networks, or influence diagrams, or (in Polish) cause-effect networks (terms frequently used interchangeably) are quite popular for expressing causal relations under multiple variable setting both for deterministic and non-deterministic (e.g. stochastic) relationships in domains like statistics, philosophy, artificial intelligence.

Various expert systems, dealing with uncertain data and knowledge, possess knowledge representation in terms of a belief network (e.g. knowledge base of the MUNIM system, ALARM network [1] etc.). A number of efficient algorithms for propagation of uncertainty within belief networks and their derivatives have been developed, e.g. [2].

## 2 Causal Inference Algorithm

Hidden (latent) variables are source of trouble both for identification of causal relationships (well-known confounding effects) and for construction of a belief network (ill-recognized direction of causal influence may lead to assumption of

independence of variables not present in the real distribution). Hence much research has been devoted to construction of models with hidden variables. It is a trivial task to construct a belief network with hidden variables correctly reflecting the measured joint distribution. One can consider a single hidden variable upon which all the measurables depend on. But such a model would neither meet the requirements put on belief network (space saving representation of distribution, efficient computation of marginals and conditionals) nor those for causal networks (prediction capability under control of some variables). Therefore, criteria like minimal latent model [3] or maximally informative partially oriented path graph [4] have been proposed. As the IC algorithm for learning minimal latent model [3] is known to be wrong, let us consider the CI algorithm from [4].

In [4] the concept of including path graph is introduced and studied. Given a directed acyclic graph  $G$  with the set of hidden nodes  $V_h$  and visible nodes  $V_s$ , representing a causal network CN, an including path between nodes  $A$  and  $B$  belonging to  $V_s$  is a path in the graph  $G$  such that the only visible nodes (except for  $A$  and  $B$ ) on the path are those where edges of the path meet head-to-head and there exists a directed path in  $G$  from such a node to either  $A$  or  $B$ . An including path graph for  $G$  is such a graph over  $V_s$  in which if nodes  $A$  and  $B$  are connected by an including path in  $G$  ingoing into  $A$  and  $B$ , then  $A$  and  $B$  are connected by a bidirectional edge  $A < - > B$ . Otherwise if they are connected by an including path in  $G$  outgoing from  $A$  and ingoing into  $B$  then  $A$  and  $B$  are connected by an unidirectional edge  $A - > B$ . As the set  $V_h$  is generally unknown, the including path graph (IPG) for  $G$  is the best we can ever know about  $G$ . However, given an empirical distribution (a sample), though we may be able to detect presence/absence of edges from IPG, we may fail to decide uniquely orientation of all edges in IPG.

Therefore, the concept of a partial including path graph was considered in [4]. A partially oriented including path graph contains the following types of edges: unidirectional  $A - > B$ , bidirectional  $A < - > B$ , partially oriented  $A o - > B$  and non-oriented  $A o - o B$ , as well as some local constraint information  $A * - * B * - * C$  meaning that edges between  $A$  and  $B$  and between  $B$  and  $C$  cannot meet head to head at  $B$ . (Subsequently an asterisk (\*) means any orientation of an edge end: e.g.  $A * - > B$  means either  $A - > B$  or  $A o - > B$  or  $A < - > B$ ). A partial including path graph (PIPG) would be maximally informative if all definite edge orientations in it (e.g.  $A - * B$  or  $A < - * B$  at  $A$ ) would be shared by all candidate IPG for the given sample and vice versa (shared definite orientations in candidate IPG also present in maximally informative PIPG), the same should hold for local constraints.

Recovery of the maximally informative PIPG is considered in [4] as too complex and a less ambitious algorithm CI has been developed therein producing a PIPG where only a subset of edge end orientations of the maximally informative PIPG are recovered. Authors of CI claim such an output to be still useful when considering direct and indirect causal influence among visible variables as well as some prediction tasks.

### 3 From CI Output to Belief Network

From the point of view of BN construction the output of CI is not satisfactory. Though it indicates necessity of inclusion of hidden variables at some places (as a parentless parent of nodes at ends of bidirectional edges), it leaves open the question of necessity of inclusion of other latent variables into the model (e.g. as parent of nodes at ends of undecided edges  $Ao - oB$  or partially decided edges  $Ao - > B$ ). Though we can clearly include as many hidden variables as we want and the model may be as correct as ever, however such procedure would impose unnecessary space and computation time burden for applications exploiting the belief network (probability distributions of hidden variables are to be saved, and the space required for saving conditional distribution at a visible node grows about exponentially with the number of ingoing edges). On the other hand, incorrect orientation of edges left unoriented may result in introducing independencies not really present in the data, so that belief network would not reflect the true underlying probability distribution. Therefore, it is of vital importance to answer the question if the output of CI is sufficient to construct a belief network, if further hidden variables need to be included, and how to construct this BN.

As an answer to this question we propose the following algorithm and the accompanying theorem. First we introduce the notion of legally removable node.

In a partially oriented including path graph  $\pi$ , a node A is called *legally removable* iff there exists no local constraint information  $B * - * A * - * C$  for any nodes B and C and there exists no edge of the form  $A * - > B$  for any node B.

#### CI-to-BN Algorithm

Input: Result of the CI algorithm (a partial including path graph)

Output: A belief network

A) Accept unidirectional and bidirectional edges obtained from CI.

B) Orient every edge  $Ao - > B$  as  $A - > B$ .

C) Copy the partially oriented including path graph  $\pi$  onto  $\pi'$ .

Repeat:

In  $\pi'$  identify a legally removable node A. Remove it from  $\pi'$  together with every edge  $A * - * B$  and every constraint with A involved in it. Whenever an edge  $Ao - oB$  is removed from  $\pi'$ , orient edge  $Ao - oB$  in  $\pi$  as  $A < - B$ .  
Until no more node is left in  $\pi'$ .

D) Understand every bidirectional edge  $A < - > B$  as indicator of parentless hidden variable  $H_{AB}$  being parent of exactly A and B  $A < - H_{AB} - > B$  with no edge between A and B.

End of CI-to-BN

**Theorem i** By the CI-to-BN algorithm, a belief network can always be obtained.  
(ii) The obtained belief network keeps all the dependencies and independencies of the original underlying including path graph.

Please notice that no other hidden variables are introduced than those indicated by CI. Notice also that the step C) is non-deterministic: at a given moment more than one legally removable node may exist. A change in order of node removal may lead to change of orientation of some edges of the resulting BN. Hence a whole family of BN compatible with the CI output and not introducing additional hidden variables is generated.

## 4 Discussion, Summary and Outlook

Within this paper an algorithm of recovery of belief network structure from data has been presented. It relies essentially on exploitation of the result of the known CI algorithm of Spirtes, Glymour and Scheines [4]. The edges of partial including path graph, not oriented by CI, are oriented to form a directed acyclic graph. We claim that such an orientation of edges always exists without necessity of adding auxiliary hidden variables, and that this BN captures all dependencies and independencies of the intrinsic underlying including path graph.

The algorithm presented will provide with belief networks with provably minimal number of parentless latent variables with two unconnected children so that the computational and spatial complexity of the resulting belief network will be as small as possible. Also the precise location of hidden variables to be included will be provided - contrary to the proposal of [1] sect.3.2.2. where most probably location of hidden variables is identified. It will (non-deterministically) yield a whole statistically indistinguishable family of such belief networks.

The CI-to-BN algorithm will suffer from very same shortcomings as the CI algorithm, that is it is tractable only for a small number of edges ( $< 10$ ). It will be interesting task to examine the possibility of such an adaptation of the Fast CI algorithm [4]. It may not be trivial as the product of CI differs from that of FCI. It is worth trying as FCI can realistically handle networks with 30 and more variables. Another path of research would be to elaborate a version of CI-to-BN assuming only with a restricted number of variables participating in d-separation, which would also bind the exponential explosion of search space.

## References

1. Cooper G.F., Herskovits E.: *A Bayesian method for the induction of probabilistic networks from data*, Machine Learning 9 (1992), 309-347.
2. Pearl J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo CA, 1988
3. Pearl J., Verma T.: *A theory of inferred causation*, [in:] Principles of Knowledge Representation and Reasoning, Proc. of the Second International Conference, Cambridge, Massachusetts, April 22-25, 1991, Allen J., Fikes R., Sandewell E. Eds., San Mateo CA: Morgan Kaufmann, 441-452.
4. Spirtes P., Glymour C., Scheines R.: *Causation, Prediction and Search*, Lecture Notes in Statistics 81, Springer-Verlag, 1993.
5. Verma T., Pearl J.: *Equivalence and synthesis of causal models*, Proc. 6th Conference on Uncertainty in AI, 220-227, 1990.