ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

A constraint-based approach to incorporate prior knowledge in causal models

G. Borboudakis[1,2], S. Triantafillou[1,2], V. Lagani[2] and I. Tsamardinos[1,2]

1- Computer Science Department, University of Crete and

2- Institute of Computer Science, FORTH , Hellas

Abstract. In this paper we address the problem of incorporating prior knowledge, in the form of causal relations, in causal models. Prior approaches mostly consider knowledge about the presence or absence of edges in the model. We use the formalism of Maximal Ancestral Graphs (MAGs) and adapt cSAT+ to solve this problem, an algorithm for reasoning with datasets defined over different variable sets.

1 Introduction

Scientific studies and experiments produce a wealth of knowledge. Often this knowledge is in the form that a quantity A is associated/not associated with B, or if the study is experimental, that A is causing/not causing B. Presumably, this type of knowledge could be helpful when performing a causal analysis on a different dataset. Existing attempts to incorporate prior knowledge when learning causal models assume *knowledge about the existence or absence of edges in the model* (see [1] for a detailed review of existing methods). The edges in most graphical models denote associations or causal relations that are not mediated by any other observed variable in the model, i.e., a direct causal relation in the context of a set of variables $\mathbf{V}$. However, when an experiment suggests that X causes Y this relation refers to any variable context; it denotes a possibly indirect causal relation in the context of $\mathbf{V}$. To impose the correct semantics of an identified relation "X causes Y" one should impose the constraint on the model that there exists a directed causal path from X to Y ($X \to \cdots \to Y$).

In this paper, we examine a constraint-based approach to incorporating prior knowledge in the form typically produced by studies, and in a way that retains the semantics of the discoveries. We use the framework of Maximal Ancestral Graphs (MAGs), and representatives for MAG equivalence classes, Partially Oriented Ancestral Graphs. These models have the desired property that can represent marginal distributions with possible latent variables and therefore do not require causal sufficiency [2]. We reduce the problem to combining causal graphs from different overlapping variable sets and employ an existing algorithm (cSAT+) to solve it. The algorithm is sound and complete and will perform all possible orientations that are unique in all models consistent with both the data and the prior knowledge. We demonstrate that complicated types of prior knowledge can be used to improve the orientations of existing models, under the assumption of correctness of the models and reliability of the prior knowledge. We include a discussion of the limitation of current formalisms such as MAGs to

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

capture interesting types of causal knowledge, as well as open questions regarding this problem.

2 Background

We briefly review some background knowledge to the problem, assuming the reader's familiarity with causal modeling. We are interested in situations where latent confounders are possible, therefore we use Maximal Ancestral Graphs [3]. Maximal Ancestral graphs are graphs with two types of edges, directed $\rightarrow$ and bidirected $\leftrightarrow$. A directed edge $A \rightarrow B$ denotes that A is a causal ancestor of B and more over, the statistical dependency between the two variables is "direct", i.e., it does not disappear conditioned on any other subset of variables in the model. A bidirected edge connects two variables if the two are correlated but none is an ancestor of the other (the two have a latent common cause). Given a MAG over a set of variables, the independencies that hold among the model variables can be obtained by the criterion of *m-separation*. Every missing edge in a MAG corresponds to a conditional independence , i.e if X is not adjacent to Y in MAG $\mathcal{M}$ over variables $\mathbf{O}$, X is independent of Y condition to a set $\mathbf{Z} \subseteq O \setminus \{X, Y\}$. MAGs are closed under marginalization: If an independence model is represented by a MAG, any marginal distribution of the model can also be represented by a MAG.

However, an independence model does not uniquely define a MAG. It defines an Equivalence Class of MAGs that share the same edges and some orientations. Such MAGs are called *Markov Equivalent* and can be represented by a graph called Partially Oriented Ancestral Graph (PAG). PAGs have three types of endpoints: Arrowheads "$>$", tails "-" and circles "$\circ$". A PAG has the same adjacencies as any member of the equivalence class, and every non-circle endpoint is invariant in any member of the equivalence class. Circle endpoints correspond to uncertainties; the definitions of paths are extended with the prefix *possible* to denote that there is a configuration of the uncertainties in the path rendering the path ancestral, inducing or m-connecting. FCI [2] is an asymptotically correct algorithm which outputs a PAG over a set of variables $\mathbf{V}$ when given access to an independence model over $\mathbf{V}$.

The causal SAT+ (cSAT+) algorithm, which in this paper we modify and use to incorporate prior knowledge, is an algorithm designed to combine PAGs over overlapping sets of variables. First, it assumes that a single causal mechanism gives rise to the marginal PAGs $\{\mathcal{P}_i\}_{i=1}^K$ observed. Each possible edge and edge-endpoint of this unknown MAG is represented with a propositional variable. The observed dependencies and independencies in all PAGs are converted to constraints among these propositional variables. The result is an instance of the SAT problem. Each and every solution to the SAT problem corresponds to a MAG $\mathcal{M}$ consistent with all input PAGs $\{\mathcal{P}_i\}_{i=1}^K$. The independence models implied by the PAGs are assumed to be correct. The MAGs that are solutions to the problem form an equivalence class of statistically indistinguishable models.

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

3 Problem definition

In this section we describe the types of prior knowledge we would like to incorporate in our models, based on our view of how scientific knowledge is produced and published.

1. X, Y are associated ($X \text{ - - } Y$) / X, Y are not associated ($X \text{ -/- } Y$)
2. X causes Y ($X \dashrightarrow Y$) / X does not cause Y ($X \not\dashrightarrow Y$)

Knowledge of the type $X \text{ - - } Y$ and $X \text{ -/- } Y$ is usually produced by typical observational studies, where a set of variables are measured on a population and an association is found or not (e.g., a doctor trying to identify risk factors for a disease). Knowledge of the type $X \dashrightarrow Y$ or $X \not\dashrightarrow Y$ stems from experimental studies (Randomized Control Trials), or causal analysis of observational data (e.g. with Causal Networks). Naturally, knowledge may also be derived from domain experts.

Notice that, $X \dashrightarrow Y$ does not semantically correspond to an edge $X \rightarrow Y$ in a causal model (Causal Bayesian Network or MAG). The former means that X is causing Y without reference to a specific variable context; the latter means that X is causing Y *directly* (i.e., no other variable is mediating the causation) in the specific context of all other variables in the model. Hence, a direct edge $X \rightarrow Y$ in a model, may disappear if more variables are added to the model making the causal relation indirect. *The correct interpretation is that $X \dashrightarrow Y$ ($X \not\dashrightarrow Y$) implies that a directed path from X to Y should (should not) exist in any causal model involving the two variables.* Similarly, $X \text{ - - } Y$ ($X \text{ -/- } Y$) implies that an m-connecting path exists (does not exist) in the model.

In this paper we attempt to incorporate such knowledge into a given model in the form of a PAG $\mathcal{P}$ over variables $\mathbf{V}$. As a first step towards this direction in this paper, we assume that the PAG contains no errors (i.e., we have perfect knowledge of conditional independencies among the variables in $\mathcal{P}$) and reliability of the prior knowledge. Given this assumption, the PAG already perfectly encodes the dependencies and independencies among variables in $\mathbf{V}$ and thus knowledge provided in the form $X \text{ - - } Y$ or $X \text{ -/- } Y$ can only re-enforce existing knowledge or conflict with it. We thus now focus on incorporating the second type of prior knowledge.

Let us first consider how the skeleton of $\mathcal{P}$ may be affected when we consider knowledge $\mathcal{K}$ of the form $X \dashrightarrow Y$ or $X \not\dashrightarrow Y$. As already mentioned the PAG already encodes the conditional and unconditional dependencies and independencies among the variables in $\mathbf{V}$. Edges in a PAG are removed only due to identified (conditional or unconditional) independencies. Relations $X \dashrightarrow Y$ and $X \not\dashrightarrow Y$ do not provide new independencies. Thus, considering knowledge $\mathcal{K}$ will not affect the skeleton of the PAG $\mathcal{P}$. Now, let us turn our attention to the end-point markings (tails "-", arrowheads "$>$", and circles "$\circ$") of the edges in $\mathcal{P}$. Given that we assume no structural errors in inducing $\mathcal{P}$, knowledge $\mathcal{K}$ may only reduce our structural uncertainty of the model. Thus, it may orient certain end-points and change them from "$\circ$" to either tails "-" or arrowheads

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

">"; however, $\mathcal{K}$ may not change already oriented end-points. Finally, we are interested in making all possible orientations and removing as many uncertainties ("∘" marks) as possible:

Problem 1. *Given a PAG $\mathcal{P}$ over variables $\mathbf{V}$ and a set of prior knowledge $\{\mathcal{K}\}_{i=1}^{M}$, where K_i is of the form $X \dashrightarrow Y$ or $X \not\dashrightarrow Y$, $X, Y \in \mathbf{V}$ induce a PAG $\mathcal{P}'$, consistent with both $\mathcal{P}$ and $\mathcal{K}$ and a minimal number of "∘" markings.*

4 Proposed Solution using cSAT+

To solve Problem 1, we transform the prior knowledge into graph path constraints, using the semantics of Ancestral Graphs and logical constraints. The proposed algorithm is shown in Algorithm Modified cSAT+, which is a modification of the more general cSAT+ algorithm [4] for this special case. The algorithm converts the problem to a SAT problem.

First, the algorithm defines the propositional variables $edge(X, Y)$ for every pair X and Y, indicating the presence or absence of an edge between X and Y; it also defines variables $arrowhead(X, Y)$ denoting that an edge between X and Y is directed from X to Y. Recall that, the truth assignment to the *edge* variables and the already oriented *arrowhead* variables is fixed, thus the decision variables in our problem is the set of *arrowhead* predicates corresponding to "∘" markings in the input PAG. We also define the derived predicate $ancestor(X, Y)$ denoting that there is a directed path from X to Y to facilitate notation; the details of the definition are in [5]. The truth status of *ancestor* is determined by the primitive variables *edge* and *arrowhead*. We can now express prior knowledge as boolean constraints with the following rules:

1. (Rule 1) If $K_i = X \dashrightarrow Y$

$$(edge(X, Y) \Rightarrow arrowhead(X, Y) \wedge \neg arrowhead(Y, X)) \wedge \\ (\neg edge(X, Y) \Rightarrow ancestor(X, Y))$$

2. (Rule 2) If $K_i = X \not\dashrightarrow Y$

$$(\neg ancestor(X, Y))$$

The SAT instance (Φ_c) is initialized with constraints created by the `PagConstraints` function. This function converts the PAG to constraints on the propositional variables: existence/absence of edges, initial edge orientations, definite non colliders, every edge has to have an arrow, no directed/ almost directed cycles. In Line 3, for each piece of prior knowledge $X \dashrightarrow Y$ or $X \not\dashrightarrow Y$ a constraint is appended to the current SAT formula.

The generated constraints include all the graph path constraints imposed by both the PAG and the prior knowledge. In the final step, we query the possible configurations of every unoriented endpoint, making all necessary orientations. Extra orientations will be done only if they are present in all possible model configurations resulting from the prior knowledge. If this is not the case, additional

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Function Modified cSAT+($\mathcal{P}, \{\mathcal{K}_i\}_{i=1}^{M}$)

$\Phi_c \leftarrow$ `PagConstraints(`$\mathcal{P}$`)`;
for *all* $\mathcal{K}_i$ **do**
  $\Phi_c \leftarrow \Phi_c \wedge rule1$ or $\Phi_c \leftarrow \Phi_c \wedge rule2$, depending on the type of $\mathcal{K}_i$;
for *every unoriented endpoint* $X * \cdots \circ Y$ *in* $\mathcal{P}$ **do**
  if $\Phi_c \wedge arrowhead(X, Y)$ *is not satisfiable* **then**
    Orient X out of Y
  else if $\Phi_c \wedge \neg arrowhead(X, Y)$ *is not satisfiable* **then**
    Orient X into Y
return $\mathcal{P}$;

knowledge cannot be incorporated in the PAG in a way that readily illustrates the information. An example for a simple scenario is shown in Figure 1.

PAGs are focused on representing independence models (sets of conditional independencies among the variables), instead of focusing on representing the causal relations. In some cases, it is impossible to semantically correctly represent both using PAGs: one has to either represent independencies that correspond to the m-separation criterion, or direct causal relations. For example, consider the causal structure $A \leftarrow B \leftarrow C$, A and B having a hidden common cause L. The PAG representing the structure when measuring only A, B, C is $A \circ\!-\!\circ B \circ\!-\!\circ C$, $A \circ\!-\!\circ C$. The edge $A \circ\!-\!\circ C$ exists because it cannot be removed conditioning on B only (it opens the path $C \rightarrow B \leftarrow L \rightarrow A$ because B is a collider on that path). Now assume that prior knowledge is available dictating that in the context of variables A, B, C there is no direct causal relation between variables A and C. One cannot incorporate this information in the PAG over variables A, B, C without altering the semantics of the graph in terms of independencies. Therefore, PAGs are not able to fully capture the existence/absence of causal relations, so other graphical models have to be defined to include cases like the above one.

5 Conclusions

We present a sound and complete algorithm that, assuming lack of statistical errors, incorporates causal prior knowledge over pairs of variables in a PAG over an overset of these pairs. The output of the algorithm is still a PAG, in terms of implied conditional independencies, but may represent configurations inconsistent with the prior knowledge. In addition, the complexity of the problem is not proved, and therefore algorithms of lower complexity may be possible. However, the algorithm introduces prior knowledge in a novel manner, as it does not constraint the graph structure to contain specific causal arcs. Instead, prior knowledge is included in the form of existence or absence of causal paths, which in our opinion corresponds to the natural way of knowledge production in sci-

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

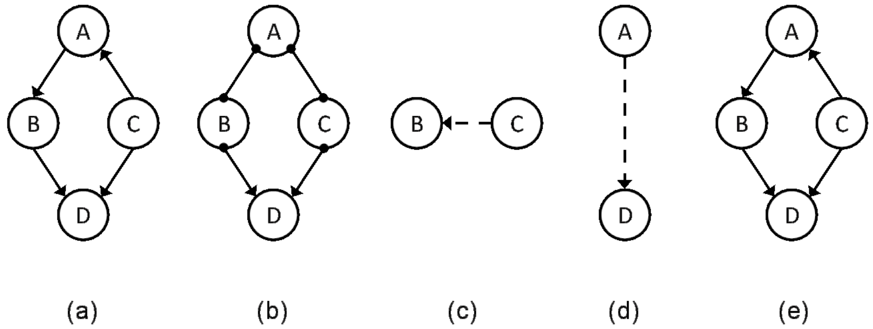

Fig. 1: **An example of the modified cSAT+ algorithm.** (a) The underlying causal structure. (b) The PAG representing the structure. (c) Prior knowledge: C is causing B. (d) Prior Knowledge: A is causing D. (e) The resulting graph by running Modified cSAT+. The dashed edges in (c) and (d) show that the causal relation is not necessarily direct, i.e. there is a causing path. First, there has to be a causal path from C to B. The only possible directed path is $C \circ\!-\!\circ A \circ\!-\!\circ B$. The result is $C \rightarrow A \rightarrow B$. To satisfy the second constraint $A \circ\!-\!\circ B \circ\!\rightarrow D$ has to be oriented as $A \rightarrow B \rightarrow D$. Finally, $C \circ\!\rightarrow D$ has to be changed to $C \rightarrow D$, because the former would result in an almost directed cycle, which is forbidden in MAGs.

entific studies. In addition, the SAT instance producec by the algorithm can be updated with novel information when available, or augmented with queries to answer specific questions of interest.

Acknowledgement: This research was partially funded by ICS, FORTH.

References

[1] N. Angelopoulos and J. Cussens. Bayesian learning of bayesian networks with informative priors. *Ann. Math. Artif. Intell.*, 54(1-3):53–98, 2008.

[2] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* MIT Press, Cambridge, MA, 2nd edition, 2000.

[3] Th. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30(4):962–1030, 2002.

[4] S. Triantafillou, I. Tsamardinos, and I. Tollis. Learning causal structure from overlapping variable sets. *Journal of Machine Learning Workshop and Conference Proceedings: AI and Statistics*, 9:860–867, 2010.

[5] S. Triantafillou. Learning causal structure from overlapping variable sets. Master's thesis, University of Crete, Heraklion, February 2010.