Learning Causal Graphs with Constraints

Murat Kocaoglu[1], **Alexandros G. Dimakis**[2], **Sriram Vishwanath**[3]
Department of Electrical and Computer Engineering
The University of Texas at Austin, USA
[1]mkocaoglu@utexas.edu
[2]dimakis@austin.utexas.edu,[3] sriram@ece.utexas.edu

Abstract

We consider causal graph learning problems with constraints. Learning is done through interventions, and the objective is to design an optimal intervention set with respect to some metric of performance or cost. We are interested in the following questions: (a) What if each intervention is limited in the number of variables it can involve? (b) What if some of the variables cannot be intervened on? (c) What if each variable has an associated cost of being intervened on? We mathematically formulate these questions. We propose a construction for (a), which is provably at least as good as the previous state of the art. We provide a full characterization for (b), explain the associated graph theory problem for (c), identify easy instances, and propose a greedy solution for the other instances. All our algorithms are non-adaptive, i.e., our interventions are designed before the outcomes of experiments are revealed.

1 Introduction

The probabilistic interpretation of causality using directed graphs was pioneered by J. Pearl in the 90s. This resulted in the widespread use of graph theoretic tools for understanding and discovering causal systems. For example, observational data from a causal system can be encoded as a set of learnable edges of a corresponding causal graph. An intervention on a set of variables corresponds to randomizing and enforcing these variables to take certain values. This corresponds to performing an experiment and collecting new data from which we can learn the directions of the edges on the cut separating the intervened and non-intervened variables. Unfortunately, for many real problems there are constraints on which variables we can perform interventions. This can be due to ethical constraints, cost or experimental complexity, or for some systems it may be simply impossible to enforce desired values for some of the variables.

In this paper, we study such problems of learning causal graphs with interventions under constraints. Specifically, we study the following problems:

Problem 1: What is the optimal intervention policy to learn a causal graph using the minimum number of interventions when each intervention is limited in size?

Problem 2: What is the optimal intervention policy to learn a causal graph using the minimum number of interventions when it is impossible to intervene on some variables?

Problem 3: What is the optimal intervention policy that minimizes the total intervention cost, when each variable has an associated cost of being intervened on?

Each of these problems is quite relevant in practice. The first problem captures the fact that experiments that contain a large number of variables are harder to perform. The second problem is important since some of the variables in a causal system may not be intervened on due to physical

29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

limitations or ethical considerations. The third problem is relevant since the experimental cost of enforcing values to different variables may vary in different applications.

In this paper we make progress on these three questions. For the problem of finding the minimum number of interventions subject to a size constraint, we propose a new technique based on replicating a small separating system across color classes. For the second problem we propose a full characterization for the optimal intervention policy. For the third problem we propose a polynomial time greedy algorithm and also identify a set of cases for which there is an exact polynomial time solution. Note that this is a summary of our work. For the proofs and an expanded background section, see [1].

2 Notation and Terminology

2.1 Observational and Interventional Learning of Causal Graphs

A directed acyclic graph $D = (V, E_0)$ on vertex set V with edge set E_0 is a causal graph if it captures the causal relationship between variables according to the Pearl's framework [8]. An undirected graph $G = (V, E)$ is the skeleton of causal graph $D = (V, E_0)$ if there is an undirected edge $(u, v) \in E$ if and only if there is a directed edge, either $(u, v) \in E_0$ or $(v, u) \in E_0$.

The causal relations within Pearl's framework can be captured through structural equation modeling with independent errors: Each variable is a deterministic function of its parent set in causal graph D and some unobserved exogenous variable. Under this model, one can show that the joint distribution of the set of variables fits into the Bayesian network D. Thus, one can use conditional independence tests in order to learn the underlying causal graph. Since Bayesian networks are not unique, one can only learn certain edges in the causal graph (See [1] for more details).

The gold standard of learning a causal graph is through experiments (also called interventions). See [8] and [9] for details. An intervention on a set S of variables yields the causal edges between S and S^c. Interventional learning can be used to complement the observational learning: One can use observational data to learn as many causal edges between variables as possible and use interventions in order to identify the remaining edges. It is known that the remaining undirected connected component after using the observational data is a chordal graph [4]. Thus, without loss of generality, we assume that the skeletons of the causal graphs we work with are chordal. This is fortunate, since chordal graphs are perfect and hence coloring is easy.

2.2 Separating Systems and Interventions

Separating systems are important constructions for interventional learning. They are used to design interventions to learn a causal graph given the skeleton [5],[6]. Given a ground set S, a separating system on S is a set of subsets $\mathcal{I} = \{I_1, I_2, \ldots\}$ that satisfies the following: For every pair of elements $u, v \in S$, there is a set I_i which contains exactly one from the pair, either u or v but not both. In other words, the set of subsets *separates* every pair of elements. A graph separating system for $G = (V, E)$ is a set of subsets of the ground set V, which satisfies the separating property for every pair of elements which form an edge.

A graph separating system on skeleton G can be used to construct a set of interventions to learn the causal graph: Simply intervene on the sets of variables in the subsets of the graph separating system. Since each edge is separated at least once, underlying causal graph can be learned.

3 Summary of Contributions

3.1 Graph Separating Systems with Constraints

Given an abstract constraint $\mathcal{C}$, we define a $(G, \mathcal{C})$ separating system as a graph separating system where each subset of the graph separating system satisfies the constraint $\mathcal{C}$. For example, if the objective is to find smallest number of interventions with a size bound on each intervention, one should find the graph separating system of minimum size, with the constraint that each subset size is at most k. We simply call these separating sytems as (G, k) separating systems.

Graph separating systems without constraints are used by [5] to design optimal set of interventions when the size of experiments is not bounded. We first show that the same correspondance between intervention designs and graph separating systems persists, even under constraints. Formally we show the following:

Theorem 1. *Consider an undirected graph G. A set of interventions $\mathcal{I}$ with constraint $\mathcal{C}$ learns every causal graph D with skeleton G if and only if $\mathcal{I}$ is a $(G, \mathcal{C})$ separating system.*

It can be shown that any graph separating system corresponds to a *not necessarily optimal* proper graph coloring. Thus, to find "optimal" set of interventions given some constraint, one can resort to the corresponding constrained graph coloring problem.

3.2 Optimal Interventions with Non-Intervenable Nodes

Consider a graph skeleton $G = (V, E)$. Let $S \subset V$ be a set of nodes we cannot intervene on. The following lemma characterizes when the underlying graph can be learned:

Lemma 1. *Let $G = (V, E)$ be an undirected chordal graph. Let $S \subset V$ be the set of non-intervenable variables. If S is not an independent set in G, then there are causal graphs with skeleton G that cannot be fully learned with any set of interventions.*

Thus if the set of nodes do not form an independent set in G, no intervention can learn every underlying causal graph D with skeleton G. Thus, we assume that S is an independent set. Then, we have the following theorem:

Theorem 2. *Consider an undirected chordal skeleton $G = (V, E)$ of a causal graph D. Let χ be the chromatic number of G. Let the set of non-intervenable nodes $S \subset V$ be an independent set of G. Then,*

a) $\lceil \log(\chi) \rceil$ interventions are necessary and sufficient to learn any causal graph D, if there exists a χ coloring of G where S is monochromatic.

b) $\lceil \log(\chi + 1) \rceil$ interventions are necessary and sufficient to learn any causal graph D, if there does not exist a χ coloring of G where S is monochromatic.

This theorem fully characterizes when any causal graph is learnable with $\lceil \log(\chi) \rceil$ experiments given skeleton G. It also shows that even when we need more interventions, $\lceil \log(\chi + 1) \rceil$ experiments are always sufficient (See [1] for construction and the proof).

Fortunately for us, there is an algorithm for perfect graphs that runs in polynomial time, which identifies if the given graph can be colored with m colors such that S is monochromatic, and outputs such a coloring if so [7].

3.3 Designs with Bounded Intervention Size

For the problem of designing bounded sized interventions, we construct a graph separating system with a size constraint on the subsets. As noted, we call this a (G, k) graph separating system, where k is the maximum size of the subsets. We have a construction, whose performance is quantified as follows:

Proposition 1. *There exists a (G, k) separating system of size $\lceil \frac{\chi\Delta}{k} \rceil \frac{\log \chi}{\log \lceil \chi\Delta/k \rceil}$.*

Authors in [6] used an (n, k) separating system to find a graph separating system where intervention size is bounded by k. This corresponds to ignoring the non-edges, and separating every pair of vertices. For an (n, k) separating system, there are constructions of size $\lceil \frac{n}{k} \rceil \frac{\log \chi}{\log \lceil \frac{n}{k} \rceil}$ [10]. Notice that, if we use the coloring that assigns a different color to every vertex, our bound reduces to this expression. Hence, over all proper colorings of the graph, this construction achieves at least as good as an (n, k) separating system, which is the best known result for this problem. Thus we have the following corollary:

Corollary 1. *Given an undirected graph G and a bound k on the subset size, we can construct (G, k) separating system using construction above, by searching over all proper colorings of G. There is always a coloring that achieves the size of an (n, k) separating system. Thus, this construction is at least as good as an (n, k) separating system with respect to the size of the separating system.*

3.4 Minimum Cost Intervention Design

For the problem of finding the cost minimizing set of interventions, given a cost of intervening per node, we can use a linear integer program using the standard coloring formulation. However, this formulation allows an optimal algorithm only for certain subgraphs, characterized in the following theorem:

Theorem 3. *Let $G = (V, E)$ be a tree graph or a clique tree, and $w : V \to \mathbb{R}$ be a cost function on its vertices. Then there is an algorithm that can find the optimum set of m interventions that can learn any causal graph D with skeleton G that runs in polynomial time.*

The proof follows from the fact that the polytope corresponding to the convex hull of all feasible points is an efficiently describable polytope for these specific graph classes, as shown in [2].

Next, we give a special case of this problem, which admits a simple solution without restricting the graph class. Assume that $m = \chi$, i.e., we are allowed to perform χ interventions. Then the problem has a simple solution, as characterized by the following theorem:

Theorem 4. *Let $G = (V, E)$ be a chordal graph, and $w : V \to \mathbb{R}$ be a cost function on its vertices. Then the optimal set of interventions with minimum total cost, that can learn any causal graph D with skeleton G is given by $\mathcal{I} = \{I_i\}_{i \in [\chi]}$, where $\{I_i\}_{i\in[\chi]}$ is any χ coloring of the graph $G_{V\setminus S} = \overline{(V\setminus S, E)}$, where S is the maximum weighted independent set of G.*

In other words, color the vertex induced subgraph obtained by removing the maximum weighted independent set S and intervene on each color class individually. The remaining graph can always be colored by at most χ colors, i.e., the chromatic number of G. Since optimum coloring and maximum weighted independent set can be found in polynomial time for chordal graphs, $\mathcal{I}$ can be constructed in polynomial time.

Another special case is when G is uniquely 2^m- colorable, if m is the maximum number of interventions we are allowed to use. Then there is only a single coloring up to permutations of colors. Hence the costs of color classes are fixed. Now we can simply sort the color classes in the order of decreasing cost, and assign row vectors of M to these color classes in the order of increasing number of 1s. This assures that the total cost of interventions is minimized.

3.4.1 A greedy algorithm

In this section, we present a greedy algorithm for the minimum cost intervention design problem, which can be run in polynomial time for interval graphs. Our approach uses the fact that finding the maximum weight $k-$colorable subgraph for interval graphs is in P due to [11].

Algorithm 1 Greedy Intervention Design for Total Cost Minimization

Input: A chordal graph G, maximum number of interventions m, cost w_i assigned to each vertex i.
$r = n, t = 0, G_1 = \overline{(V_1, E)}, V_1 = V$.
while $r > 0$ **do**
    Find maximum (weighted) $\binom{m}{t}-$colorable induced subgraph S_t
    Assign all weight$-t$ binary vectors of length m as rows of $M(S_t, :)$ to different color classes.
    $G_{t+1} = \overline{(V_{t+1}, E)}, V_{t+1} = V_t \setminus S_t$: G_{t+1} is the induced subgraph on the uncolored nodes.
    $r \leftarrow |V_{t+1}|$: r is the number of uncolored nodes.
    $t \leftarrow t + 1$
end while
return M.

In the algorithm, $\overline{(V_{t+1}, E)}$ represents the induced subgraph on the node set V_{t+1}. The algorithm returns a matrix M, which is a graph separating system for G. The idea is to choose as large a weight as possible for binary vectors with the smallest weight. This assures that the sets of nodes with large total cost are intervened fewer number of times. Steps for finding the maximum weighted $\binom{m}{t}$ colorable subgraph are given in [11], and it has polynomial runtime for interval graphs. When the graph is chordal (not necessarily interval), finding the maximum (weighted) $k-$colorable subgraph is $NP-$hard when k is not a constant, and is in P when k is a constant.

For chordal graphs, we can modify the algorithm as follows: Instead of iteratively coloring until no uncolored node remains, run the algorithm for a steps, for some constant a. Then, the maximum (weighted) $k-$ colorable subgraph can be found in polynomial time. This will make sure $\sum_{i=0}^{a} \binom{m}{i}$ colors are used. The uncolored graph at this step G_{m+1} is perfect since perfectness is hereditary. As the final step, color this graph with minimum number of colors and use the remaining small-weight vectors for these colors that are not yet used. This modification allows the algorithm to run in polynomial time even for chordal graphs.

References

[1] Extended version:. `https://sites.google.com/a/utexas.edu/mkocaoglu/InterventionDesign.pdf`. Accessed: 2016 - 10 - 31.

[2] Diego Delle Donne and Javier Marenco. Polyhedral studies of vertex coloring problems: The standard formulation. *Discrete Optimization*, 21:1–13, 2016.

[3] Frederick Eberhardt. *Causation and Intervention (Ph.D. Thesis)*, 2007.

[4] Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13(1):2409–2464, 2012.

[5] Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal networks from interventional data. In *Proceedings of Sixth European Workshop on Probabilistic Graphical Models*, 2012.

[6] Antti Hyttinen, Frederick Eberhardt, and Patrik Hoyer. Experiment selection for causal discovery. *Journal of Machine Learning Research*, 14:3041–3071, 2013.

[7] Jan Kratochvil and Andras Sebo. Coloring precolored perfect graphs. *Journal of Graph Theory, John Wiley & Sons*, pages 208–215, 1995.

[8] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.

[9] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. A Bradford Book, 2001.

[10] Ingo Wegener. On separating systems whose elements are sets of at most k elements. *Discrete Mathematics*, 28(2):219–222, 1979.

[11] Mihalis Yannakakis and Fanica Gavril. The maximum k-colorable subgraph problem for chordal graphs. *Information Processing Letters*, 24:133–137, 1987.