Protecting causal effects

May 12, 2016

Differential privacy has primarily focused on protecting individuals but there is increasing interest in problems relating to hiding certain aggregate properties of a dataset whilst preserving the ability to use it for a specified purpose. In this problem we consider if we can release a dataset for predictive purposes but discourage the inference of causal conclusions about relationships between the covariates.

Consider a scenario under which the causal graph generating a dataset is considered known - but may contain unmeasured variables.

Assume we add noise to the dataset via a single point crossover process (see paper). The goal is to prevent reliable estimation of causal effects without effecting our ability to predict a particular target variable. Some example questions:

- Under what circumstances (graph structures) it is possible to disrupt inference of a particular causal effect always provided there is a direct causal relationship between the exposure variable X and outcome variable Y, ie $P(Y|do(X)) \neq P(Y)$
- How can we maxiumully disrupt this inference (ie is there a cut that adds more noise for a given amount of shuffling of the data) data dependent
- What about if we want to disrupt a specific set of causal inference questions
- What about if we want to disrupt as many causal queries as possible.

1 Problem Statement and Definitions

- Let G be a causal directed acyclic graph over vertices V and edges E.
- Let $S = \{(X_1, Y_1), (X_2, Y_2), ..., (X_m, Y_m) : X_i, Y_i \in V, X_i \neq Y_i\}$ be a set of pairs of vertices in G.
- Each pair of vertices (X_i, Y_i) in S represents the causal query $P(Y_i|do(X_i))$
- A set of variables $Z \subset V$ is an adjustment for a query (X, Y) if $X, Y \notin Z$ and $P(Y|do(X)) = \sum_Z P(Y|X, Z)P(Z)$
- An adjustment for a query is minimal if it does not contain another adjustment as a proper subset.
- The Single Point Crossover Process partitions the variables V into two disjoint sets, $\mathcal{P}_1$ and $\mathcal{P}_2$ such that $\mathcal{P}_1 \cup \mathcal{P}_2 = V$. Let $\mathcal{P}_X$ denote the partition containing the variable X.

1.1 Discouraging inference of a set of causal queries via adjustment

Adjusting for nuisance or confounding variables, either by matching or regression, has been a key technique in inferring causal effects at least as far back as . The widely used back-door criterion **?** clarifies which variables it is appropriate to condition on in order to achieve unbiased estimates of causal effects.

Lemma 1. *Where X and Y are singleton variables, Z is a minimal covariate adjustment for identifying the causal effect of X on Y if and only if Z is a minimal set satisfying the backdoor-criterion relative to X and Y Textor and Liskiewicz (2012).*

The existence of a set covariate adjustment is sufficient but not necessary for identifiability. A causal query can also be identified

We will consider non-parametric point identifiability as described by **?**.

The do-calculus provides a complete framework to determine if and how a causal effect may be identified from observational data. The framework assumes that the data is generated by a directed acyclic graph (DAG) and that the structure of the graph is known (although some variables may not be observed).

is a sufficient criterion for identification of causal effects via matching. More recently it has been generalized to provide necessary and sufficient critera for identification via matching for DAGs , and MAGs PAGs .

More recently, **?** developed the do-calculus, which provides a complete framework for determining if a causal query can be identified from observational data.

Definition 2. Identifiability. A query $P(Y|do(X))$ is identifiable in a causal graph G if it can be computed uniquely from any positive probability distribution over the observed variables - that is, we can compute an expression for $P(Y|do(X))$ that contains only factors of the observed distribution.

1. $P(Y|do(X)) = P(Y)$.
2. $P(Y|do(X)) = \sum_Z P(Y|X,Z)P(Z)$
3. $P(Y|do(X))$ is identifiable but not via (1) or (2)
4. $P(Y|do(X))$ is not identifiable

In this paper we focus on jamming queries that fall into case 2. Case 1 cannot be jammed as the SPCP does not modify marginal distributions. However, it is also unlikely to be sensitive as it holds only if there is no casual relationship between X and Y. In case 4, the causal effect cannot be identified from this data set so there is no further work required to jam it. Case 3 is complex and rarely used in empirical work. We leave it for future work.

Lemma 3. *Let Z be a set of variables that is a minimal adjustment for the causal query (X,Y). Inferring $P(Y|do(X))$ via adjusting for Z is jammed if $Y \cup Z \notin \mathcal{P}_X$. In other words, if at least one of the variables in $X \cup Y \cup Z$ is in a different a different partition to the others.*

Note: this holds only if Z is a minimal adjustment. If Z is not minimal then we can let $Z = Z' \cup U$ where Z' is a minimal adjustment. In this case, we can still estimate the causal effect of X on Y by conditioning on Z'.

This stems from the fact that the expression for calculating $P(Y|do(X))$ via adjusting for Z contains the factor $P(Y|X,Z)$. I'm assuming the single point crossover process gives us this result. We need to a clear definition of 'jammed' in terms of the crossover-process.

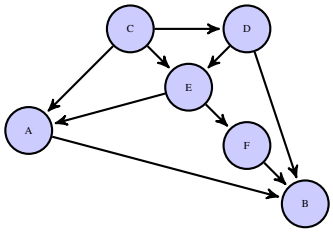
Lemma 4. *Let $\boldsymbol{Z}$ be the set of minimal adjustments for (X,Y). The causal query (X,Y) is jammed if inferring $P(Y|do(X))$ via Z is jammed for $\forall Z \in \boldsymbol{Z}$*

1.1.1 An algorithm for finding a cut that renders jams a set of causal queries S

1. For each query $(X_i, Y_i) \in S$, find the set of minimal adjustments $\boldsymbol{Z}_i$
2. Construct the set of sets $Q = \{X_i \cup Y_i \cup Z_{ik} : 1 \leq i \leq m, 0 \leq k \leq |\boldsymbol{Z}_i|\}$
3. Find a partition such that $(Q_i \nsubseteq \mathcal{P}_1) \wedge (Q_i \nsubseteq \mathcal{P}_2) \qquad \forall Q_i \in Q$

For example in figure 1, with $S = \{(A,B),(E,B),(C,B),(E,A)\}$

Figure 1



$S_1 = (A, B)$, $\boldsymbol{Z}_1 = \{\{C, E\}, \{D, E\}\{D, F\}\}$
$S_2 = (E, B)$, $\boldsymbol{Z}_2 = \{\{C, D\}\}$
$S_3 = (C, B)$, $\boldsymbol{Z}_3 = \{\}$
$S_4 = (E, A)$, $\boldsymbol{Z}_4 = \{C\}$
$Q = \{\{A, B, C, E\}, \{A, B, D, E\}, \{A, B, D, F\}, \{E, B, C, D\}, \{C, B\}, \{E, A, C\}\}$

We could jam S with any partition of the form $\{A, B, ...\}, \{C, E, F, ...\}$.$S_2, S_3$ and S_4 are jammed because their X and Y variables are in different partitions. S_1 is jammed because, although A and B are in the same partition, for each of the three minimal adjustment sets, at least one variable is not in $\mathcal{P}_A$. D could be placed in either partition. This is not the only solution, $\{A, C, ...\}, \{B, E, ...\}$ is another. Note than in any solution B and C must be in different partitions as this is the only way to jam S_3.

The number of minimal adjustments for a given query can grow exponentially with the number of nodes in the graph. There is an algorithm to enumerate them that requires $O(n^3)$ per adjustment Textor and Liskiewicz (2012) .

1.1.2 Some things we can say

- If $\boldsymbol{X} = \{X_1...X_m\}$ and $\boldsymbol{Y} = \{Y_1...Y_m\}$ are disjoint sets, then any partition with $\boldsymbol{X} \subseteq \mathcal{P}_1$ and $\boldsymbol{Y} \subseteq \mathcal{P}_2$ will jam S. This holds trivially if $|S|$ = 1.
- The hardest case to jam is when $\boldsymbol{Z}_i = \forall i$
- If $|S| = 2$ we can always jam S as either $\boldsymbol{X}$ and $\boldsymbol{Y}$ will be disjoint or a variable A will appear in both S_1 and S_2, in which case we can let $\mathcal{P}_1 = \{A\}$.

1.2 Further work

We have considered only single variable interventions and outcomes. We believe the results generalize in a straightforwardly to the case where, for each causal query, X and Y are sets with a minor assumption that there are no causal paths from one intervention variable X to another.

Although overwhelmingly used in practice, adjustment is not necessary for causal identifiability. A given causal graph may also be identifiable via the front door criterion or via an expression returned by the identify algorithm .

What is the psuedo code to jam all mechanisms? We need an algorithm to produce all 'minimal' expressions. Then each expression will contain a set of factors. At least one of those factors must be jammed for every expression.

Note: A given query may be identifiable via multiple different expressions. As we must jam all of them, it is not sufficient to have an algorithm that returns *an* expression (if one exists). We are not aware of any such algorithms for the general identifiability problem.

Prove that a single query can still always be jammed by separating X and Y. Approach - the expression is a result of the do calculus and nessesarily contains a term with both X and Y together unless...

References

Textor, J. and Liskiewicz, M. (2012). Adjustment criteria in causal diagrams: An algorithmic perspective. *arXiv preprint arXiv:1202.3764*.