Protecting causal effects

May 8, 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

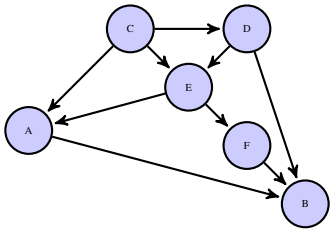
- Let G be a causal directed acyclic graph over vertices V and edges E.
- Let $S = \{(X_1, Y_1), (X_2, Y_2), ..., (X_m, Y_m) : X_i, Y_i \in V, X_i \neq Y_i\}$ be a set of pairs of vertices in G.
- Each pair of vertices (X_i, Y_i) in S represents the causal query $P(Y_i|do(X_i))$
- A set of variables $Z \subset V$ 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

Lemma 1. *One of the following holds.*

1. $P(Y|do(X)) = P(Y)$

Figure 1



2. *There is no set of variables that satisfy the back-door criterion for identifying* $P(Y|do(X))$
3. *There is at least one set of variables* Z *that are a minimal adjustment for* $P(Y|do(X))$*,* Z *may be the empty set.*

For the remainder we assume that all queries we wish to jam are of case 3, because in case 1, there is no causal effect of X on Y to discover and in case 2, the causal effect of X on Y can already not be inferred from the data and no way in which we partition the data will render it identifiable.

Lemma 2. *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* $Y \cup Z \not\subset \mathcal{P}_X$*. 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.*

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 3. *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 $(Q_i \not\subseteq \mathcal{P}_1) \wedge (Q_i \not\subseteq \mathcal{P}_2) \qquad \forall Q_i \in Q$

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

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

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 $\mathcal{P}_1 = \{A\}$.

References

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