Causal Inference with Graphical Neural Networks

Your Name

April 9, 2024

1 Introduction

2 do Operator

The **do** operator is a way to represent interventions in a causal model. It is a way to represent the effect of an intervention on a variable. As an example, consider the following model involving smoking.

If a person's fingernails (N) have turned yellow, this implies a higher probability that they are a heavy smoker (S) and hence have a higher probability of developing lung cancer (C). But, simply dyeing a persons fingernails yellow does not impact their probability of developing lung cancer.

So, in terms of **do** calculus, we can denote the process of setting a variable N to have a value *yellow* by $\mathbf{do}(N = \textit{yellow})$. We note that

$$P(C \mid N = \textit{yellow}) \neq P(C \mid \mathbf{do}(N = \textit{yellow})).$$

With this in mind, we now define the **do** operator.

Theorem 2.1 ([?]) *In a causal diagram Γ with nodes $X_1, \ldots, X_n$ and joint distribution $P(X_1, \ldots, X_n)$, the result of doing $X_i = x_i$ on the joint distribution is*

$$P(X_1, \ldots, X_n \mid \boldsymbol{do}(X_i = x_i)) = \frac{P(x_1, \ldots, x_n)}{P(x_i \mid \operatorname{par}(x_i))} = \prod_{j \neq i} P(x_j \mid \operatorname{par}(x_j)).$$

In this, we have $\operatorname{par}(x_i)$ represent values of the parent nodes of $\operatorname{PAR}(X_i)$ of X_i in Γ. The probabilities on the right hand side of the above equation are what we call *preintervention*. This means they use the original probabilities from the original model before doing $X_i = x_i$.

It is important to note that the above equation is how we calculate the probability of several events happening given one event has happened. What if we want to get the probability of a single event happening, given we do a single event? That leads to the following corollary.

Corollary 2.1.1 *If X and Y are random variables in a causal diagram Γ and* $\mathrm{PAR}(X)$ *are the parents of X, then*

$$P(y \mid \boldsymbol{do}(x)) = \sum_{\mathrm{par}} \frac{P(x,\, y,\, \mathrm{par})}{P(x \mid \mathrm{par})},$$

where the sum runs over all values par *that the variables* $\mathrm{PAR}(X)$ *can take. If X has no parents, then*

$$P(y \mid \boldsymbol{do}(x)) = \frac{P(x,\, y)}{P(x)} = P(y \mid x).$$

Let us now consider a basic example to see how this works. Consider the following causal diagram in Figure 1.

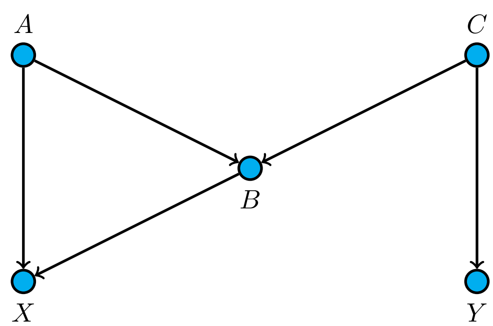

Figure 1: Basic causal diagram. Note it is in the form of a directed acyclic graph (DAG).

In this diagram, we can see that A and C are both parents of B. So, for any values of x and b, Corollary 2.1.1 tells us that

$$P(X = x \,|\, \mathbf{do}(B = b)) = \sum_{\mathrm{par}(b)} \frac{P(x,\, b,\, \mathrm{par}(b))}{P(b \,|\, \mathrm{par}(b))}$$

which, written out, is

$$\sum_{\mathrm{par}(b)} \frac{P(x,\, b,\, \mathrm{par}(b))}{P(b \,|\, \mathrm{par}(b))} = \sum_{a} \sum_{c} \frac{P(X = x\,,\, A = a\,,\, B = b\,,\, C = c)}{P(B = b \mid A = a,\, C = c)}.$$

By dependence of nodes only on their parents and the rules of probability, this turns into

$$\sum_a \sum_c \frac{P(X = x \mid A = a\,,\, B = b)P(B = b \mid A = a,\, C = c)P(A = a)P(C = c)}{P(B = b \mid A = a,\, C = c)},$$

which simplifies to

$$\sum_a \sum_c P(X = x \mid A = a\,,\, B = b)P(A = a)P(C = c).$$

Since there is only one instance where we are considering the probability with respect to c, we can simplify this to

$$\sum_a P(X = x \mid A = a\,,\, B = b)P(A = a),$$

which is our final answer.

While this introduction to the **do** operator might feel a bit abstract, it is the foundation of all current research in causal inference.

3 Data

For our project, we used the LUCAS0 dataset [**?**], which is a toy data set generated artificially by causal Bayesian networks with binary variables. The LUCAS0 dataset is a DAG with 11 nodes and 2000 training samples, where the DAG is represented as

4 Background

5 Methodology

6 Experiments

6.1 Dataset Description

6.2 Experimental Setup

6.3 Results

7 Discussion

8 Conclusion

Acknowledgments

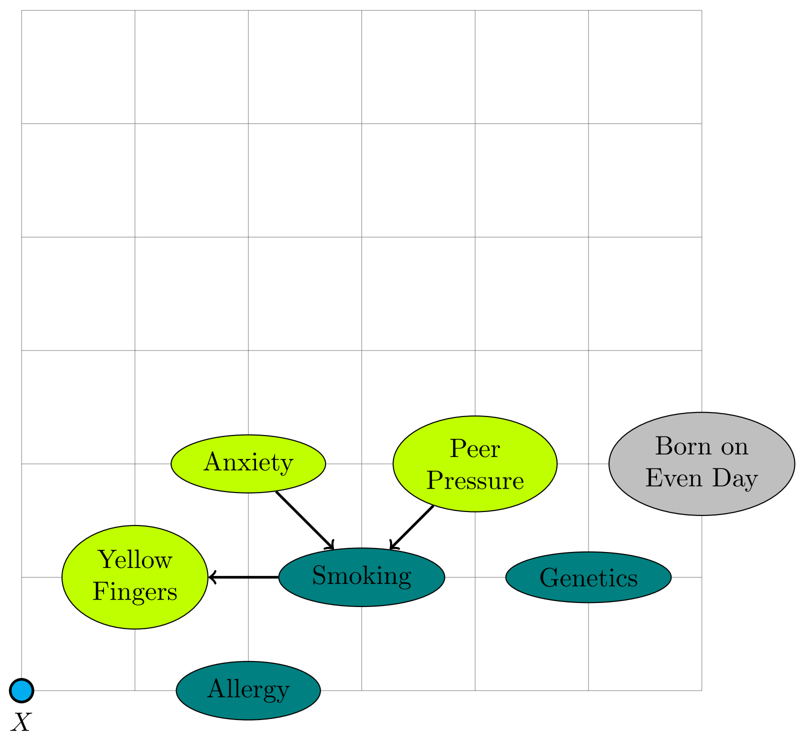

Figure 2: Basic causal diagram. Note it is in the form of a directed acyclic graph (DAG).