Causal Sum-Product Networks

Abstract

Lorem ipsum...

1 INTRODUCTION

Lorem ipsum...

2 PRELIMINARIES

We begin with sum-product networks (SPN) by Poon & Domingos [2011] which represent a special type of probabilistic model that allow for a variety of exact and efficient inference routines. Following that we provide a primer to important notions from causality literature to be used in subsequent sections.

2.1 SUM-PRODUCT NETWORKS

Members of the family of probabilistic circuits[1] that satisfy the properties decomposability and smoothness are known as sum-product networks[2]. SPN are graphs consisting of three types of nodes (product, sum and leaf nodes) whose structure and parameterization can be efficiently learned from data to allow for efficient modelling of joint probability distributions. [...]

2.2 CAUSALITY

[...] A Structural Causal Model (SCM) as defined by Peters et al. [2017] (see Definition 6.2) is specified as $\mathfrak{C} := (\mathbf{S}, P_{\mathbf{N}})$ where $P_{\mathbf{N}}$ is a product distribution over noise variables and $\mathbf{S}$ is defined to be a set of d structural equations

$$X_i := f_i(pa(X_i), N_i), \quad \text{where} i = 1, \ldots, d \tag{1}$$

with $pa(X_i)$ representing the parents of X_i in graph $G(\mathfrak{C})$. An intervention on a SCM $\mathfrak{C}$ as defined in (1) occurs when (multiple) structural equations are being replaced through new non-parametric functions $\hat{f}(\widehat{pa(X_i)}, \hat{N}_i)$ thus effectively creating an alternate SCM $\hat{\mathfrak{C}}$. Interventions are referred to as *imperfect* if $\widehat{pa(X_i)} = pa(X_i)$ and as *atomic* if $\hat{f} = a$ for $a \in \mathbb{R}$. [...] An important property of interventions often referred to as "modularity" or "autonomy"[3] states that interventions are fundamentally of local nature, formally

$$p^{\mathfrak{C}}(x_i \mid pa(x_i)) = p^{\hat{\mathfrak{C}}}(x_i \mid pa(x_i)) \tag{2}$$

where the intervention of $\hat{\mathfrak{C}}$ occured on variable X_k opposed to X_i. The statement in (2) suggests that mechanisms remain invariant to changes in other mechanisms which implies that only information about the effective changes induced by the intervention need to be compensated for. An important consequence of autonomy is the truncated factorization formula

$$p(V) = \prod_{i \notin S} p(X_i \mid pa(X_i)) \tag{3}$$

derived by Pearl [2009] which suggests that an intervention S introduces an independence of an intervened node X_i to its causal parents $pa(X_i)$. [...]

[1] An umbrella term introduced by Van den Broeck [2019].

[2] For a formal treatise of decomposability and smoothness, Section 2 of Peharz et al. [2020] should be considered.

[3] See Section 6.6. in Peters et al. [2017].

3 CAUSAL SUM-PRODUCT NETWORKS

In this section, we identify how to approximate arbitrary interventional distributions by extending SPN such that they can adhere to causal quantities.

3.1 PROBLEM STATEMENT

Peters et al. [2017] (Figure 1.4) motivate the necessity of causality for adequate generalizability of predictive models. Furthermore, we argue that ignoring causal change in a system, i.e., the change of structural equation(s) underlying the system, can lead to significant performance decrease. Therefore, it is important to account for distributional changes [...]

3.2 APPROACH

We extend the idea of conditional parameterization for SPN presented in Shao et al. [2019] by conditioning on the Pearl's *do*-operator [2009] while predicting the complete set of observed variables. Mathematically, we estimate the conditionals $p(V \mid do(S=s))$ by learning a non-parametric function approximator $f_{\boldsymbol{\theta}}$ (e.g. neural network), which subsumes knowledge on the causal structure $G(\mathfrak{C})$ and the intervention $do(S=s)$, to predict the parameters of a learned SPN $p_{f_{\boldsymbol{\theta}}}(V)$. The joint model $M=(f_{\boldsymbol{\theta}}, p_{f_{\boldsymbol{\theta}}})$ we denote as causal sum-product network (Causal SPN).

Proposition 1 (Expressivity)**.** *A Causal SPN $M=(f_{\boldsymbol{\theta}}, p_{f_{\boldsymbol{\theta}}})$ can identify any interventional distribution permitted by SCM $\mathfrak{C}$.*

Proof. The Causal SPN M approximates a joint distribution $\hat{p}(V)$ that is consistent with the factorization $p(V)=\Pi_{i \notin S} p(X_i \mid pa(X_i))$ of the (intervened) causal graph $G(\mathfrak{C}^{do(S=s)})$. The truncated factorization that follows Pearl's *do*-calculus states that the statistical quantity $p(V)$ identifies the causal quantity $p(V \mid do(S=s))$. Thus $\hat{p}(V) \approx p(V) = p(V \mid do(S=s))$. □

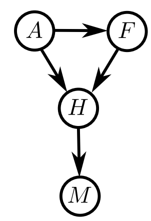
Figure 1 visualizes the approach applied to the Causal Health SCM being presented in Section 4.

4 EMPIRICAL EVALUATIONS

In this section, we discuss results on an implementation of causal sum-product networks. The implementation is provided [...]. We evaluate against a newly curated synthetic dataset before benchmarking against existing methods on competitive datasets.

4.1 SYNTHETIC DATASET: CAUSAL HEALTH

To demonstrate the expressivity of causal sum-product networks in modelling arbitrary interventional distributions, we curate a causal dataset based on the SCM presented in Figure 2. In the following, the SCM will be denoted by $\mathfrak{C}$ and non-empty interventions on the SCM by $\mathfrak{C}^{do(S=s)}$ where S are the variables to be intervened upon and s are their respective instantiations.



$$A = U(18, 80)$$
$$F = \mathbf{1}_F(A) \vee B(\tfrac{1}{2})$$
$$H = \mathbf{1}_H(A) + F + B(\tfrac{1}{2})$$
$$M = \tfrac{1}{2}H + B(\tfrac{1}{2})$$

Figure 2: **Causal Health SCM.** The causal graph (left) and the structural equations (right) for the data generating process are presented. The names for the depicted variables are: A is Age, F are Food Habits, H is Health, and M is Mobility. The noise distributions are either Uniform $U(\cdot,\cdot)$ or Bernoulli $B(\cdot)$. The indicator functions are defined as $\mathbf{1}_F(A=a) := 1$ if $a \geq 40$, 0 if $a < 40$ and $\mathbf{1}_H(A=a) := 0$ if $a \geq 60$, 1.5 if $a \geq 30$, 3 if $a < 30$ respectively.

The SCM $\mathfrak{C}$ describes the causal relations of an individual's health and mobility attributes with respect to their age and nutrition. The Causal Health dataset does not impose assumptions over the type of random variables or functional domains of the structural equations[4] which additionally constraints a learned model to adapt flexibly. A causal sum-product network is being learned using a set of mixed-distribution samples generated from simulating the Causal Health SCM for different interventions[5], mathematically $\mathrm{D} = \{do(S_i = s_i), X_{\mathfrak{C}}\}_i$ with $X_{\mathfrak{C}} \in \mathbb{R}^{1000 \times 4}$. We assess the performance of our learned model in adapting to previously unseen interventions using the Jensen-Shannon divergence $D_{\text{JSD}}(p||q) = \frac{1}{2}D_{\text{KL}}(p||m) + \frac{1}{2}D_{\text{KL}}(q||m)$ based on the asymmetric Kullback-Leibler divergence $D_{\text{KL}}(p||q) = \sum_{x \in X} p(x) \log_2(\frac{p(x)}{q(x)})$[6] and $m = \frac{1}{2}(p+q)$. Two distributions p, q are equal if and only if $D_{\text{JSD}} = 0$. Table 1 shows the performance of our method [...]

[4] Assumptions on the functional form of structural equations are considered crucial for identifiabilty of causal quantities (see Table 7.1 in Peters et al. [2017])

[5] The observational case is considered to be equivalent to an intervention on the empty set.

[6] Using $\log_2(\cdot)$ bounds the Jensen-Shannon divergence such that $D_{\text{JSD}}(p||q) \in [0, 1]$.

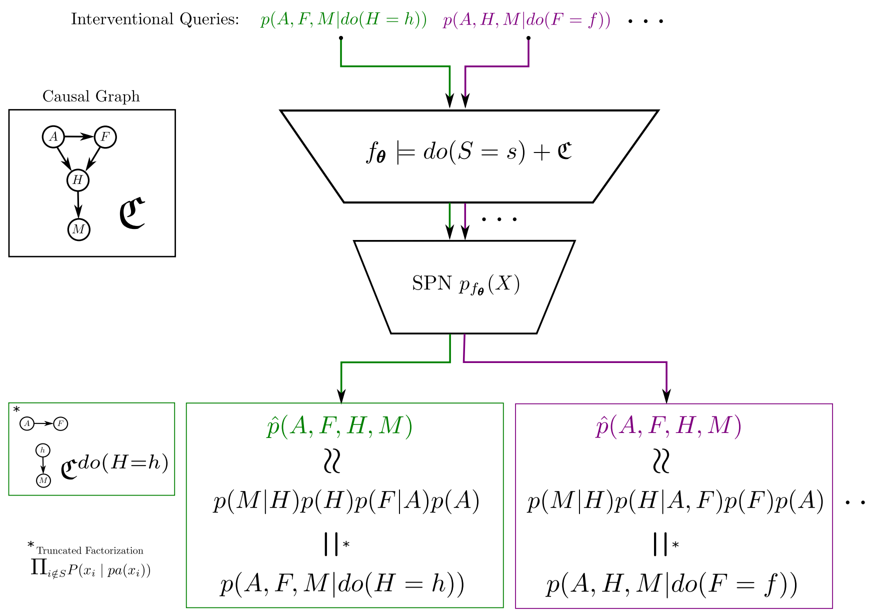


Figure 1: **Causal SPN Example.** Given interventional queries $p(V \mid do(S = s))$ on a provided dataset (in this example, the Causal Health SCM from Figure 2 is being considered thus $V = \{A, F, H, M\}$), the Causal SPN $M = (f_{\boldsymbol{\theta}}, p_{f_{\boldsymbol{\theta}}})$ predicts a joint distribution $\hat{p}(V)$ that aligns with the truncated factorization of the intervened graph $p(V) = \Pi_{i \notin S} p(X_i \mid pa(X_i))$ which identifies the targeted causal quantity $p(V \mid do(S = s))$.

p	q	$D_{\mathbf{JSD}}(p\|\|q)$
$do(\emptyset)$	$do(H = U(H))$	0.74
Causal SPN	$do(H = U(H))$	**0.24**

Table 1: **Causal SPN Performance**. Measured using Jensen-Shannon divergence $D_{\text{JSD}}(\cdot, \cdot)$. Distributions with $do(\cdot)$ are generated with the Causal Health SCM $\mathfrak{C}$.

5 CONCLUSIONS

Lorem ipsum...

References

Judea Pearl. *Causality*. Cambridge university press, 2009.

Robert Peharz, Steven Lang, Antonio Vergari, Karl Stelzner, Alejandro Molina, Martin Trapp, Guy Van den Broeck, Kristian Kersting, and Zoubin Ghahramani. Einsum networks: Fast and scalable learning of tractable probabilistic circuits. *arXiv preprint arXiv:2004.06231*, 2020.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference*. The MIT Press, 2017.

Hoifung Poon and Pedro Domingos. Sum-product networks: A new deep architecture. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 689–690. IEEE, 2011.

Xiaoting Shao, Alejandro Molina, Antonio Vergari, Karl Stelzner, Robert Peharz, Thomas Liebig, and Kristian Kersting. Conditional sum-product networks: Imposing structure on deep probabilistic architectures. *arXiv preprint arXiv:1905.08550*, 2019.

G Van den Broeck, N Di Mauro, and A Vergari. Tractable probabilistic models: Representations, algorithms, learning, and applications. *Tutorial at UAI*, 2019.