Deep Structural Causal Models for Tractable Counterfactual Inference

Nick Pawlowski[=], Daniel C. Castro[=], and Ben Glocker

BioMedIA Imperial College London

NEURAL INFORMATION PROCESSING SYSTEMS

Overview

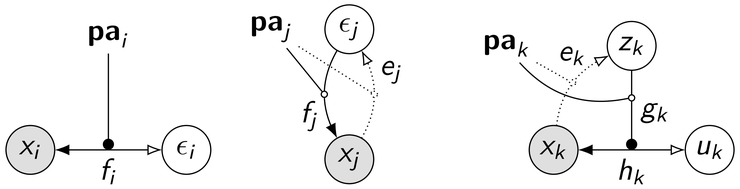
- Unified framework for structural causal models (SCMs) with deep generative models, enabling tractable, plausible counterfactuals for high-dimensional data
- Case studies demonstrate how to apply deep structural causal models and perform counterfactual inference
- Limitation: assumes no unobserved confounders and fully observed data during training and counterfactual inference
- Further work should investigate learning dynamics of deep mechanisms and their identifiability

Observations to Counterfactuals

Consider a SCM with assignments $\{x_k := f_k(\epsilon_k; \mathbf{pa}_k)\}_{k=1}^K$ and noise distribution $P(\boldsymbol{\epsilon}) = \prod_{k=1}^K P(\epsilon_k)$. We can perform counterfactual inference with 3 steps:

1. **Abduction:** infer noise from observations, $P(\boldsymbol{\epsilon}|\mathbf{x})$
2. **Action:** perform intervention, e.g. $\text{do}(x_k := \widetilde{x}_k)$
3. **Prediction:** plug noise back into SCM, $x_j = f_j(\epsilon_j; \widetilde{\mathbf{pa}}_j)$

Deep Mechanisms



'Flow-based' 'GAN-based' 'VAE-based'

Abduction:

$$\epsilon = f^{-1}(x; \mathbf{pa}) \quad \Big| \quad \epsilon \approx e(x; \mathbf{pa}) \quad \Big| \quad \begin{aligned} z^s &\sim Q(z|e(x; \mathbf{pa})) \\ u^s &= h^{-1}(x; g(z^s; \mathbf{pa}); \mathbf{pa}) \end{aligned}$$

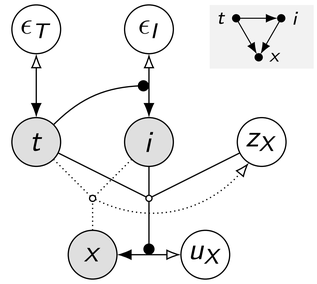
Action: intervene on upstream variables $(\mathbf{pa} \mapsto \widetilde{\mathbf{pa}})$

Prediction:

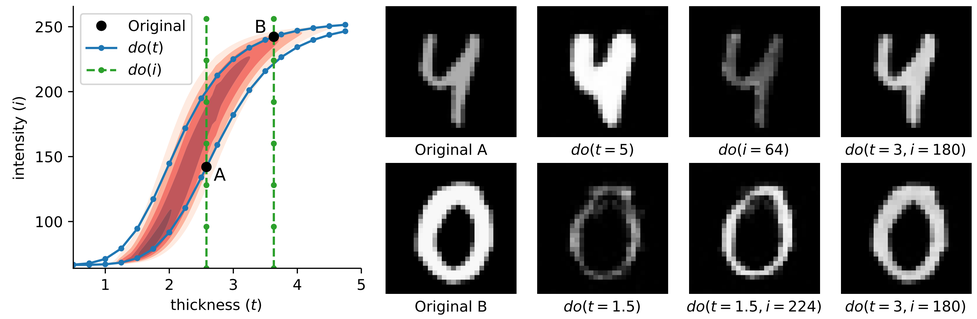
$$\widetilde{x} = f(\epsilon; \widetilde{\mathbf{pa}}) \quad \Big| \quad \widetilde{x}^s = \widetilde{h}(u^s; \widetilde{g}(z^s; \widetilde{\mathbf{pa}}), \widetilde{\mathbf{pa}})$$

Morpho-MNIST* Experiment

We train a DSCM corresponding to the true data-generating process of a synthetic dataset of thickness (t), intensity (i) and image (x):

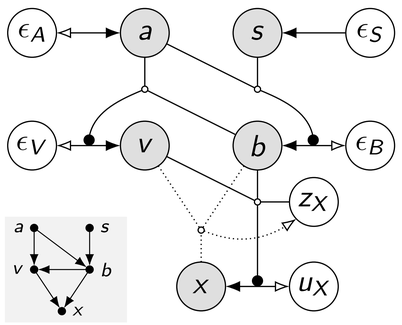


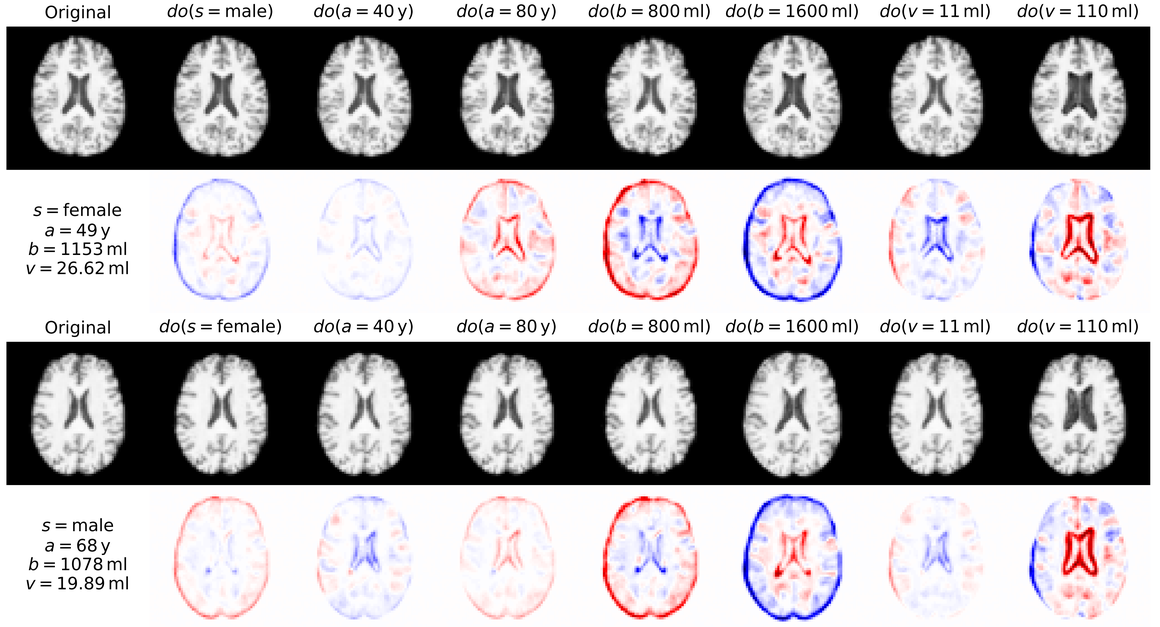
Infer counterfactuals for images A and B. The density plot shows the change of the covariates given counterfactual interventions and the images the relevant image counterfactuals:



UK Biobank Experiment

Clinically inspired DSCM using age (a), sex (s), ventricle volume (v), brain volume (b) and 2D mid-axial slices of brain MRI:





/biomedia-mira/deepscm

arXiv.org/abs/2006.06485

*DC Castro et al. (2019). Morpho-MNIST. *JMLR*, 20(178)

project mira ERC StG 2017 grant No 757173

erc European Research Council Established by the European Commission