
CausalStructCodec: Causally-aware observational and interventional data generator

Louis Hernandez^{*†1} and Matthieu Boussard¹

¹Craft AI – Craft AI, Craft AI – France

Abstract

Generative models are useful tools for data augmentation or private data sharing. In the realm of causality, Structural Causal Models (SCM) are theoretical models representing the true data-generating process. They allow for the computation of interventional and counterfactual generation. But these models are theoretical and, in practice, we do not have access to the functions of the SCM that truly generate the data. We can however have access to the causal graph, using domain knowledge. Dedicated generative models have emerged to leverage that knowledge of the causal graph to generate data, mimicking an SCM, such as MultiCVAE, CAREFL or VACA. However, we found that these models have some limitations, including computational efficiency and lack of flexibility regarding complex data structures. Therefore, we provide the CausalStructCodecs (CSC), a novel architecture, based on Codecs, designed to tackle these issues. We show that it generates interventional and observational data with at least equivalent accuracy to state-of-the-art models while taking advantage of the performance and flexibility of the Codecs architecture.

Keywords: Generative Models, Codecs, Causality, Interventions

^{*}Speaker

[†]Corresponding author: louis.hernandez@craft.ai