Abstract

Understanding cause-effect relationships improves the performance of different machine learning tasks, including prediction, recommendation, and decision-making. Causal reasoning plays a pivotal role in this process, as it uncovers the underlying cause-effect relationships within data, enabling better generalization and adaptation to dynamic changes. Consequently, causal reasoning has gained significant importance and is now applied across diverse domains.

This research explores causal reasoning considering three perspectives: the preservation of causal relations during data generation, the identification of causal sensitivity, and the learning of cause-and-effect relationships in the presence of unobserved confounders. This work emphasizes the role of generative learning in achieving these objectives.

The study first investigates the extent to which generative models preserve causal relationships during data synthesis, without being provided with the underlying causal structure. Causality preservation is evaluated under realistic settings, including completely unknown structures with observed and unobserved common causes, as well as partially known scenarios where conditioning on causally sensitive concepts is available. To quantify preservation, we introduce an evaluation framework termed inverse causal mapping, which measures the consistency between generated samples and the true cause–effect mechanisms. This study measures the significance of causally sensitive features for preserving causality during data generation. It establishes that conditioning on causally sensitive concepts correctly follows the causal directions and preserves more causality.

The second component of this research focuses on identifying causal sensitivity by determining causally sensitive features and aims to identify target-specific causal paths by applying interventional and counterfactual aspects of causality in the absence of unobserved confounders. The framework evaluates the significance of causally sensitive features in predictive tasks and quantifies their impact.

Finally, this research addresses causal structure learning when the underlying cause-effect relationships are unknown, particularly in the presence of both observed and unobserved confounders. To enhance structure learning, we introduce causal constraints into generative models along with a novel causal annealing strategy that progressively guides training toward stable and meaningful causal representations. The objective is to recover both directed and confounded relationships while strengthening the capacity of generative models to retain causal dependencies, thereby contributing to a robust framework for generative causal reasoning.

Experimental evaluations are conducted on synthetic SCM-generated data and multiple real-world datasets, including BNLearn benchmark networks, UCI datasets (e.g., Abalone), the Sachs protein signalling dataset, Morpho-MNIST, spatio-temporal human

trajectory data, and electronic health records. Comparative analysis against state-of-the-art methods demonstrates the robustness, generalizability, and effectiveness of the proposed framework across preservation, identification, and causal structure learning scenarios.

Keywords: Generative Causal Models, Counterfactual Inference, Causal Structure Learning, Latent Confounding