

Project Proposal: Inducing Causal Structure from Data

Causal Structure Learning

Inferring the causal structure of a set of variables from data is a fundamental problem in science with many applications. From determining the causal effect of a drug on patients to predicting the economy, causal structure learning is a crucial component of many scientific disciplines.

The Paper

The Paper "Learning to Induce Causal Structure" by [Ke et al.](#), a poster from ICLR 2023, proposes an approach that treats the inference process as a black box and uses a variant of the transformer architecture whose attention mechanism is structured to discover the causal structure of the data.

The Model Architecture

The model uses a transformer encoder and a transformer decoder.

The Encoder

The encoder is structured as an $(N + 1) \times (S + 1)$ lattice that receives an $N \times S$ dataset. Furthermore, the data is embedded with an MLP, and the transformer uses alternating attention. This encoder produces a summary vector.

The Decoder

The decoder uses the summary information from the encoder to generate a prediction of the Adjacency matrix of the causal graph with an autoregressive transformer.

Why this paper?

The paper is recent and tackles an interesting problem in a novel way. The empirical results recorded by the paper are impressive based on the reviewers' comments. Synthetic data can be generated with a causal Bayesian network, while there is also a lot of real-world data available. The architecture is novel and interesting while also challenging to implement. Depending on the dataset's size, the model's training cost can be kept low by using a small number of variables, or if the model is not too expensive to train, a large dataset can be used to train the model, potentially on a cloud service.

Possible Evaluations

Other work uses a score function to search the space of possible graphs and find the best one. There is also the approach of using a NN to learn the scoring function. We could compare the performance of the model to these approaches.

Possible Difficulties

Training the model will consume a lot of training data since it takes an $N \times S$ dataset as input. The model is also quite complex. However, the paper provides adequate detail on most of the model architecture.