# My Research: Causality in practice? – It ain't easy.

## Sebastian Weichwald, 2022-05-24

My research focusses on what hurdles fruitful practical applications of causal modelling and causal machine learning.

**Speech recognition, text translation, automatic video captioning, … isn't machine learning already fruitfully applied?**

The deep learning approach to machine learning has arguably revolutionised data science. Fitting expressive function approximators and high-dimensional density estimators as black-boxes to huge datasets led, for example, to impressive language and computer vision tools. Yet, in high-risk applications, such as treatment planning and therapy chatbots for depressed or suicidal patients, we cannot deploy such black-box models and risk severe failures. Instead, we need models with reliable and expectable model behaviour also under environmental shifts and interventions. For example, a vision system for autonomous driving ought to be robust against environmental shifts and to be working reliably in a white winter landscape, on a colourful autumn parkway, and between trucks with large-area natural paintings on them.

So we often require modelling abilities beyond classical probabilistic and machine learning models that predict future observations of a system under the same conditions that held true when we devised the model (observational distribution). Perhaps, causality is what we need. With a statistical *causal* model we may also reason about future observations of the system when subjected to external manipulating forces or environmental shifts (interventional distribution).

This ability is crucial for reliable predictions and data-driven decisions as the following example aims to illustrate (here, it suffices to understand the graph intuitively as a depiction of the cause-effect relationships, while its causal semantics and correspondence to an accompanying structural equation model can also be made formally precise):

$$\text{diabetes} \leftarrow \text{insulin deficiency} \rightarrow \text{blurred vision} \leftarrow \text{myopia}$$

A non-causal model may wrongly lead us to treat blurred vision (the side effect) instead of insulin deficiency (the cause) since blurred vision predicts but does not cause diabetes (the effect); likewise, the model may fail to accurately predict the incidence of diabetes in a different environment and population with prevalence of myopia (another cause of blurred vision) higher than in the training population.

Indeed, research into causality in machine learning gained traction, not least since Judea Pearl's 2011 Turing Award on causality, and research foci shifted accordingly, for example, from representation learning to causal representation learning.

**Causal machine learning to the rescue?**

The models deployed today for prediction tasks on text, image, video, and voice data are highly complex and non-linear with billions of parameters, composed of dozens of stacked neural network layers, and take days to train in the cloud. This sheer model complexity and flexibility appears to be key to the success of predictive machine learning and allows scaling to high-dimensional (low-risk) applications such as ad placement or text translation. While equipping these machine learning models with causal reasoning abilities is promising, it ain't easy.

First, in statistical causal modelling we often require restrictive assumptions for providing a statistical account of causal model properties and for characterising their identifiability from data. The gap is substantial between machine learning engineering practice and statistical causal modelling theory. On the one hand are universal function approximators, massive datasets, flexible high-dimensional density estimators, and a model zoo of often heavily overparameterised models; on the other hand are restricted model classes with limiting assumptions, for example, on the linearity of cause-effect relationships or the (non-)existence of relevant yet unobserved variables, and impossibility results that characterise the theoretical limits for learning causal structure from data. While machine learning models over thousands of variables are commonplace, for example, for high-resolution images, we struggle to scale causal structure learning algorithms beyond a couple of dozens of variables. To make progress despite this gap and to improve machine learning via causal principles, I research how to use abstract causal knowledge as weak supervision signal for improving rich machine learning models. The hope is to systematically incorporate our prior knowledge, which is often not only partial but comes in abstract form of high-level properties instead of exact causal knowledge on the model variables, into complex machine learning pipelines.

Second, in causal inference we often implicitly assume that the observable variables are meaningful causal entities. This assumption is often inapt and hurdles fruitful application. In neuroimaging, for example, it is nonsensical to apply a causal structure learning algorithm to the raw fMRI voxel values or EEG signals. Before learning the causal structure, we need to construct representations or recover latent from observed variables that are meaningful causal entities. We face similar issues when we transform variables that admit a causal model to simplify the model. For example, even if we could obtain a causal model over several billions of neurons, this model may not be pragmatically useful for a neurologist to decide on patient treatment plans. To obtain a useful comprehensible model, we need to abstract and simplify the billions-of-neurons model while retaining the essential causal information. Only by systematically deriving a manageable model on the level of brain regions from a causal model over billions of neurons do we obtain a model that is pragmatically useful for a neurologist to decide on patient treatment plans. In both cases, when we recover latent causal variables or aggregate variables to simplify a model, we easily break the variables' and model's causal semantics. Conversely, the choice of wrong variables or the inability to measure the right variables conceptually hinders the applicability of causal inference methodology. To address this hurdle, we formalised and characterised causal model transformations and abstractions that guarantee that the causal semantics stays intact when transforming the underlying variables. We hope to extend this to a notion of approximate correspondence between two causal models with an associated approximation error that can be used to learn optimal simplifications of causal models.

Third, the scarcity of real-world data with known cause-effect relationships hinders the development of causal structure learning algorithms. Instead of real-world data, benchmarks of causal structure learning algorithms commonly rely on synthetic data and presume causal additive noise models. In my research I scrutinise the practical relevance of such benchmarks and additive noise models. We have shown that without knowledge of the true data scale recent structure learning algorithms that follow a direct machine learning paradigm fail to recover causal structure. Similarly, our insights from the Causality 4 Climate NeurIPS competition 2019 highlight the (unintended) patterns in data scale of additive noise models: we demonstrated that for causal models with additive noise simple algorithms achieve performance on par with other more involved and computationally expensive causal discovery algorithms. In my research I aim to improve benchmarking of causal structure learning and to raise awareness that assumptions commonly employed in the causality literature may arguably be implausible for real-world settings and thus prevent fruitful application of the developed tools.