How to Model Causality: From Philosophy to AI

Zhuoying Li
Yuanpei College
Peking University
joy@stu.pku.edu.cn

Abstract

This essay explores the concept of causality from philosophical and mathematical standpoints, studying its development from Hume to Lewis, and onward to the use of probability in the creation of causal models. It considers the advantages and drawbacks of Rubin's Causal Model (RCM) and Structural Causal Model (SCM). Inspired by Hume's Regularity Theory of Causation (RTC), it suggests the amalgamation of the probabilistic approach with Multimodal Large Language Models (MLLMs) as an innovative solution to accurately analyze event correlations and capture complex spatiotemporal contexts, thereby unpacking causality in a more comprehensive manner.

1 Introduction

The concept of causality has consistently been a topic of intrigue, from the philosophical discourses of David Hume to the era of computers and artificial intelligence. Hume and Lewis provided a base, defining 'cause' and paving the way for modern interpretations. Two such interpretations, Rubin's Causal Model (RCM) and Structural Causal Model (SCM), have leveraged probability to simplify causality's intricate nature. However, these probabilistic models have their limitations, particularly when it comes to understanding the temporal and spatial context. Notably, they excel in identifying correlations between events but struggle with deeper aspects of causality. This shortcoming opens the door for an innovative approach, integrating these models with Multimodal Large Language Models (MLLMs).

2 What is causality: From Hume to Lewis

Before we model causality, it's important to define what is "causality". In this section, we will discuss Hume and Lewis' definition of "cause".

In [1], Hume defines "cause" in the following two ways:

- *"We may define a cause to be an object followed by another, and where all the objects similar to the first are followed by objects similar to the second..."*
- *"Or in other words where, if the first object had not been, the second never had existed."*

While it might appear that the second definition is merely a reiteration of the first, they each approach the concept of 'cause' from a distinct angle. The first definition tackles causality in a forward manner that suggests a regularity, while the second views it through a reverse, or counterfactual lens.

However, the first definition presents certain difficulties. The phrase 'followed by' is a somewhat elusive term, merely describing an observable correlation between events without accounting for the directional aspect which causality necessitates. Consider the classic scenario in which a farmer observes a correlation between the rooster crowing and the sun rising, and erroneously concludes that the rooster's crow instigates the sunrise. Additionally, it's important to understand that these occurrences may not have any direct cause-and-effect relationship. They could both potentially be the effects of a separate shared cause. Take, for instance, the correlation between the rising sales of ice cream and an increase in swimming-related fatalities. These two occurrences often coincide, but not because one causes the other. Rather, they're most likely influenced by a common factor, which is the seasonal pattern of the summer when such activities are typically higher.

Lewis, drawing inspiration from Hume's second definition, proposed the concept of *Counterfactual Dependence* [2]. Roughly, b causally depends on a if and only if b holds true in all the most similar possible worlds in which a holds true. In other words, the occurrence of one event depends on what did or did not happen in another event.

3 Use probability to model causality: RCM and SCM

In this section, we will discuss two ways to model causality and analyze their advantages and problems.

3.1 Rubin's Causal Model (RCM)

Rubin's Causal Model (RCM) is a statistical framework employed to estimate causal effects in observational or experimental data [6]. The principle idea in RCM is that to define causal effects rigorously, we must contemplate potential outcomes. Take the example of measuring the effect of aspirin on headache: For each unit (i.e., individual, group, etc.), there are two potential outcomes, corresponding to two treatments (0 or 1). Let's assume $Y_i(1)$ for treatment and $Y_i(0)$ for control, where i denotes an individual. Then the causal effect for unit i is the difference between these potential outcomes, i.e., $Y_i(1) - Y_i(0)$.

In real-life situation, however, causal inference always involves untestable assumptions about counterfactuals (i.e., what would have happened if we had done something different). For example, if we want to estimate the causal effect of an educational intervention on students' exam performance, we might compare the performance of students who received the intervention ($Y_i(1)$) to the performance of students who didn't ($Y_i(0)$). However, we can't observe both potential outcomes for a single student. Therefore, we have to make assumptions to estimate the Average Treatment Effect (ATE), which is given by $ATE = E(Y(1) - Y(0))$.

3.2 Structural Causal Model (SCM)

The Structural Causal Model (SCM) was first proposed by Judea Pearl and his team. It not only helps to determine the correlation between variables, but also to understand what changes will occur if we intervene on a system. Pearl defines a causal model as an ordered triple $\langle U, V, E \rangle$, where U is a set of exogenous variables whose values are determined by factors outside the model; V is a set of endogenous variables whose values are determined by factors within the model; and E is a set of structural equations that express the value of each endogenous variable as a function of the values of the other variables in U and V [4]. Pearl also proposed *the ladder of causation* [5], which is a three-level abstraction of causality:

- Association: statistical relationships among variables in terms of observations
- Intervention: what happens when external forces override part of the system
- Counterfactuals: an alternate version of a past event, or what would happen under different circumstances for the same experimental unit

Let's assume that you are trying to determine if smoking causes lung cancer, then the ladder of causation can be described as follows:

- Association level: Initially, you observe that individuals who smoke indeed have a higher likelihood of developing lung cancer. This is mere correlation, not causality. Mathematical representation: $P(cancer \mid smoking)$
- Intervention level: Now, implement an anti-smoking campaign as an external intervention. If the incidence of lung cancer decreases, you form a hypothesis that smoking indeed leads to lung cancer. Again, this doesn't prove causality outright but strongly suggests it. Mathematical representation: $P(cancer \mid do(smoking))$, where *do* is an operator that signals the experimental intervention (implement an anti-smoking campaign)
- Counterfactual level: You could then consider what would have happened if one of the smokers who now has lung cancer, had never taken up smoking (rewinding life scenarios). If the counterfactual

scenario suggests the person wouldn't have lung cancer, you can link smoking and lung cancer. Mathematical representation: $P(cancer \mid no\ smoking)$

3.3 Discussions

RCM fundamentally aligns with Hume's first definition of "cause", focusing more on regularity and correlations between events. As stated in Sec. 2, correlation doesn't imply causality, but causation always implies correlation. Hence, ambiguity may arise within this model, potentially muddling the causal chain of certain events.

SCM takes things a step further by incorporating a causative diagram that illustrates the causal interconnection of variables within the model. Additionally, it uses the ladder of causation principle derived from Lewis' counterfactual dependence. This allows for a more accurate representation of causality. However, the intricacies involved in drafting the correct causal diagrams often require significant effort. Moreover, SCM often oversimplifies the causal relationships. It assumes that the real-world phenomena can be perfectly represented in a mathematical model, which is often not the case.

4 Beyond probability: An inspiration from Hume's theory

To solve the problem of probabilistic method, we can look back to Hume's theory. In [1], Hume proposes his famous Regularity Theory of Causality (RTC), which can be described as follows:

- Temporal priority: cause must precede the effect in time
- Spatial contiguity: proximity in space between the cause and the effect
- Constant conjunction: multiple observations, correlation

The probabilistic approach excels in identifying correlations between events, but it falls short in capturing temporal and spatial context. This shortcoming prompts the consideration of MLLMs. It's been observed that LLMs adeptly grasp the temporal aspects of an input sequence. Recently, models infused with multimodal information have gained significant traction. A notable example is InstructBLIP [3], which can effectively process both images and text to deal with multiple tasks. Thus, integrating the probabilistic approach with MLLMs may offer a promising solution. This combination could accurately analyze event correlations while capitalizing on temporal and spatial context to unveil causality.

References

[1] David Hume. *A Treatise of Human Nature*. Oxford University Press, Oxford, 1978. revised P.H. Nidditch. 1, 3

[2] David Lewis. Causation. *Journal of Philosophy*, 70(17):556–567, 1973. doi: 10.2307/2025310. 2

[3] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *arXiv preprint arXiv:2304.08485*, 2023. 3

[4] Judea Pearl. *Causality*. Cambridge university press, 2009. 2

[5] Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, Inc., USA, 1st edition, 2018. ISBN 046509760X. 2

[6] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974. 2