Building a computational causal model

Zhiqiang Luo
Department of Automation
Tsinghua University
luo-zq21@mails.tsinghua.edu.cn

Abstract

Causality is an essential part of human cognition, enabling us to understand the underlying causal structure of the environment. Investigating causality requires the creation of computational models that represent causal structures and causal relations. In this essay, we identified three features that a computational causal model should possess, including an explainable structure, a probabilistic representation, and the ability to incorporate prior information effectively. And we present a comprehend introduction of one computational causal model: Causal Bayes Net.

1 Introduction

The concept of causality is centered on the fundamental causal relationships between events. Causality manifests itself in diverse domains, spanning from elementary physical causality, extracted during the perceptual phase [6], to intricate social causality, characterized by the presence of numerous variables [7]. Causality-based causal reasoning is a cognitive tool upon which individuals depend in their daily lives, enabling them to comprehend their surroundings and make informed decisions regarding how to interact with their environment. For instance, awareness of the causal relationship between smoking and lung cancer explains why individuals who smoke are at a higher risk of developing lung cancer. This understanding can, to a certain extent, serve as a deterrent against smoking, and prevent individuals from it.

In the context of the aforementioned example, the causal structure of smoking and lung cancer corresponds to a mental representation of a causal model. When explaining why smokers face an elevated risk of lung cancer, we engage in causal inference rooted in this mental model. The choice to reduce smoking can be regarded as a form of intervention or a manifestation of counterfactual thinking. These concepts of causal inference, intervention, and counterfactual thinking hold substantial importance in the field of causal reasoning. **Causal inference** equips individuals with the capacity to generate data that aligns with the causal model. **Intervention** provides individuals with the means to manipulate variables to achieve desired outcomes. Similar to intervention, **counterfactual thinking** follows a paradigm akin to "if A happened, B would have happened", where A did not happen in reality.

Causal reasoning equips individuals with the ability to adapt to unfamiliar environments through a process known as causal transfer. Causal transfer allows individuals to extrapolate the causal structure of a new environment by leveraging their understanding of the causal relationships in a familiar setting, facilitating a swift and efficient adaptation to the novel environment.

The development of a computational model of causal relations and causal structures is essential to the research in causality, offering indispensable mathematical tools. Griffiths and Tenenbaum [3] discussed that a computational-level analysis focuses on abstract problems and logic, meaning that the adoption of a computational model will free us from unnecessary debates about representations and implementations within the human mind. In Sec. 2, we will delve into some essential features of a computational causal model. Subsequently, in Section Sec. 3, we will provide a comprehensive introduction to the implementation of a computational causal model known as the Causal Bayes Net.

2 Essential features of a computational causal model

A computational model of causal relations and causal structures serves as a mathematical tool for research in human causality, encompassing aspects like causal inference and causal induction. In this section, we will explore three essential features of a computational causal model: explainable structure, probabilistic representation, and sensitivity to prior information.

2.1 Explainable structure

The requirement for an explainable structure in a computational model arises from its role in explaining the intricacies of the environment. In various fields of cognition, such as affordance and intuitive physics, models like neural networks may reign supreme owing to their impressive performance, even though they often lack an explainable structure. However, within the field of causality, our pursuit focuses precisely on the absent component in neural networks: the underlying causal relationships between events. This distinction clarifies why neural networks find limited utility in the domain of causality research.

Furthermore, the presence of an explainable structure within a computational model enhances its capacity for causal transfer. Edmonds et al. [1] demonstrated the limitations of model-free Reinforcement Learning (RL) models in the tasks of causal transfer. Due to their reliance on actions and rewards, model-free RL models lack explainable structures, resulting in shortcomings in generalization and causal transfer. With an explainable structure, a computational model can swiftly and effectively adapt its structure when confronted with a new environment, eliminating the need to relearn an entirely new structure from scratch.

2.2 Probabilistic representation

The notion of a probabilistic representation is introduced in opposition to a deterministic representation. A probabilistic representation means that "if A cause B, then the occurrence of A will lead to the occurrence of B with a probability of .8". On the contrary, a deterministic representation means that "if A cause B, then the occurrence of A will definitely lead to the occurrence of B" [4].

Besides performing causal inference, people learn causal models from evidence and actions, which is an inverse problem of causal inference called causal induciton [2] [3]. If we were to employ a deterministic representation in computational models, it would be challenging to discern the correct and suitable causal structure from the myriad of potential structures that align with the evidence patterns. Conversely, by embracing a probabilistic representation, we can identify the most plausible causal structure based on the posterior probabilities of these potential structures, significantly simplifying the inverse problem, causal induction [2].

2.3 Sensitivity to prior information

In the domain of causal induction, numerous potential causal structures may align to the observed evidence patterns. Discerning the correct and suitable causal structure could demand a substantial amount of data, often unattainable in practice. To address this challenge, individuals leverage prior information, thereby reducing the range of potential causal structures to a limited space of hypotheses. Similarly, a computational causal model should exhibit the sensitivity to prior information to effectively narrow down the space of hypotheses [3].

3 An implementation: Causal Bayes Net

In this section, we will present an exemplar of a computational causal model: the Causal Bayes Net (CBN), as articulated by Pearl et al. [5]. This model has garnered significant use in the field of causality research. However, it's important to recognize that the Causal Bayes Net is just one of numerous potential implementations of computational causal models, and new models may attain more impressive performance than CBN.

Within the framework of a CBN, the foundational structure for representing causal relationships is an acyclic directed graph. In this graphical model, nodes represent pertinent states, while the arrows connecting two nodes signify the causal relationship between these states, with the arrow pointing from the causal factor to the effect. Each relationship is assigned with a functional form, indicating how causal factors affect the effect. If we note a node as s, its causal factors as $pa(s)$, and the probability distribution function as P, then a functional form F can be expressed as

$$F(pa(s)) = P(s \mid pa(s))$$

.

With a foundational understanding of CBN, we can delve into the process of conducting causal inference with CBN. Causal inference within the CBN framework is essentially a top-down problem. Leveraging the probabilistic distribution of certain initial variables, we can compute the probabilistic distribution of the target variable by employing the graphical structure and functional forms defined within the model [2]. When engaging in more intricate causal inference tasks, such as interventions and counterfactual thinking, it becomes necessary to manually fix specific variables within the graph. By doing so, we eliminate the target variable's relationships with its causal factors. For example, consider the scenario where playing video games leads to staying up late, which subsequently results in lower grades. If we perform intervention on the variable "staying up late", then no matter whether we play video games, we will sleep at the same time, indicating the elimination of the relationship between "playing video games" and "staying up late". The intervention on the variable s is denoted as $do(s)$. When performing interventions or counterfactual thinking on s about a target variable t, we are actually calculating

$$P(t \mid do(s))$$

.

CBN makes it possible to perform causal induction, which means infer the graphical structure and parameters of functional forms from accessible data and actions [2]. The causal induction of CBN relies on Bayesian inference:

$$P(c \mid data) \propto P(data \mid c)P(c)$$

, where c denote a hypothesis of a CBN. Therefore, we can calculate the posterior probability of a hypothesis based on the prior probability and the likelihood of the data, and infer the most possible CBN within the hypothesis space. Edmonds et al. [1] have utilized Bayesian inference to infer the causal structures of the OpenLock problems. In their work, they learn the likelihood $P(data \mid c)$ and calculate the prior probability $P(c)$ based on the distribution of atomic structures. They planned the action based on the posterior probability and updated the distributions and parameters based on the outcomes of the action, until the model could solve the OpenLock problems.

Nonetheless, CBN faces a challenge associated with their expansive hypothesis space, often demanding substantial computational resources for Bayesian inference [3] [8] [1]. In response to this issue, Griffiths and Tenenbaum [3] introduced theory-based causal induction, a method that effectively harnesses prior environmental knowledge to simplify the causal inference process. With such prior knowledge including ontology, plausible relations, and functional forms, the size of hypothesis space is reduced to an acceptable scale.

4 Conclusion

In this essay, we have explored the field of computational models designed to represent causal structures and relations. We have highlighted key challenges within the domain of causality research, specifically in the context of causal inference and causal induction. Furthermore, we have identified and elucidated crucial features that a computational causal model should possess, including an explainable structure, a probabilistic representation, and the ability to incorporate prior information effectively. Lastly, we have introduced an exemplary instantiation of a computational causal model, the Causal Bayes Net.

References

[1] Mark Edmonds, Xiaojian Ma, Siyuan Qi, Yixin Zhu, Hongjing Lu, and Song-Chun Zhu. Theory-based causal transfer: Integrating instance-level induction and abstract-level structure learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1283–1291, 2020. 2, 3

[2] Alison Gopnik and Henry M Wellman. Reconstructing constructivism: causal models, bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, 138(6):1085, 2012. 2, 3

[3] Thomas L Griffiths and Joshua B Tenenbaum. Theory-based causal induction. *Psychological review*, 116(4):661, 2009. 1, 2, 3

[4] Christopher Hitchcock. Causal Models. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2023 edition, 2023. 2

[5] Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2):3, 2000. 2

[6] Martin Rolfs, Michael Dambacher, and Patrick Cavanagh. Visual adaptation of the perception of causality. *Current Biology*, 23(3):250–254, 2013. 1

[7] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974. 1

[8] Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, et al. Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering*, 6(3):310–345, 2020. 3