Toward a New Representation for Causation in Dynamic Systems

Denver Dash*
Intel Science & Technology Center on
Embedded Computing at
Carnegie Mellon University
Pittsburgh, PA 15213
denver.h.dash@intel.com

Mark Voortman
Decision Systems Laboratory
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA, 15260, USA
mark@voortman.name

1 Introduction

Over the past twenty years, causal modeling has been a growing discipline within the field of machine learning (ML). Inspired by the rallying cry of Pearl and others, the ML community has provided a wealth of methods for approximate and exact causal discovery and causal reasoning. Causation has a unique place in machine learning and AI for several reasons: Causal models are generative probabilistic models which seek to provide the "most parsimonious" representation for describing an *entire system*, as opposed to discriminitive classification models which are interested in predicting a fixed set of variables. This may make causal models suitable for a "general AI" agent whose scope is intended to exceed any simple classification problem. Another reason is that causal models aid explanation because they mirror the way many humans internally model the world [Sloman, 2005]. Perhaps most importantly, causal models provide a syntax for reasoning about manipulating elements of the system being modeled. In this paper we explore the limits of the state-of-the-art causal representations, we identify some practical obstacles to their use, and suggest a new representation to help remedy them.

2 Causation in Philosophy and Machine Learning

Causation in philosophy has been a highly contentious issue. Hume [1739], for example, argued that causation is neither grounded in formal reasoning nor in the physical world. Therefore, he concluded, causation is nothing more or less than a habit of mind, albeit a useful one. Russell [1913] went as far as saying that causation is "a relic of bygone age," maintaining instead the view that symmetric physical laws are the important representation. Although he later retracted from this view and expressed more favorable opinions towards causation. Needless to say the debate did not end with Russell or Hume, and causation is still a raging debate within philosophy. One of the central issues in this debate is how to define causality in a non-circular way. Modern practitioners in the social sciences, ML and AI all circumvent this issue by defining causality cleanly in the context of a model.

Formal models of causation were developed in econometrics by Simon [1952] and Wold [1954]. Wold represented causation as a set of *structural equations*, i.e., equations ordered in such a way that each variable was solved in terms of other variables. Such a system unambiguously defines a directed graph over the variables of the system. We call this line of research the *structural equation* or the *directed graph* approach to causation.

The structural equation approach is essentially the representation advocated by Spirtes et al. [1993] and Pearl [2000] who also co-identified a set of assumptions under which the structure of causal models may be recovered from observational data. This work has been very influential in machine learning and led to many extensions involving dynamic models, constraint-based and Bayesian methods as well as methods which combine observational with experimental data. Thus directed graph representations of causation is the predominant representation used in machine learning today.

*Also Adjunct Assistant Professor, Department of Biomedical Informatics, University of Pittsburgh.

The view of causation proposed by Simon [1952] was slightly different. Instead of pre-solving each variable in terms of its causal parents, a system is described as a set of fundamental *mechanisms* which corresponded to invariant laws of nature governing the system. These mechanisms by themselves are symmetric, but once they are placed in the context of a fixed set of boundary conditions, they lead to assymmetric causal relations. Simon defines the *Causal Ordering Algorithm* which, given such a completely specified system, outputs a partial ordering of the variables and thus constrains the causal graphs that can be defined over the system. We call this representation the *mechanistic* approach to causation. The mechanistic representation does not output a fully directed graph (some sets of nodes may be "coupled" together), but it is capable of handling a wider concept of manipulation than that defined over directed graphs, as we explain below.

2.1 Manipulation in Causal Models

The ability to reason about manipulating variables in a system is one of the chief features that motivates the development of a good causal representation. For example, having a detailed causal model of a particular biological subsystem can assist in understanding how different drugs might impact that subsystem. Having a detailed causal model of how cancer develops from cigarettes can lead to healthier nicotine delivery mechanisms, etc.

Manipulation in structural-equation-based causal models is especially simple, and also originated in econometrics with Strotz and Wold [1960], but was popularized in machine learning by the well-known Do operator of Pearl [2000]. In a nutshell, when the Do operator operates on a variable X in a causal graph, it effectively removes all incoming links into X. This simplicity is one of the primary benefits of structural-equation-based causal models. Stated in terms of structural equations themselves, the equation $X = f(Pa(X))$ that determines the variable X gets replaced with the simpler equation $X = X_0$.

Similarly, manipulation in mechanism-based causal models is achieved by replacing an equation with a different equation; however, since there is not a 1-1 correspondence between variables and equations, this operation does not necessarily correspond to removing edges in the causal graph. This can be useful when, for example, coincident with manipulating one variable you wish to "release" another. This is a richer concept of manipulation than allowed by the Do operator, and can lead to much different manipulated causal graphs. In fact, sometimes very non-intuitive effects on the causal graph can be obtained, such as reversed causal edges [Leijen and Druzdzel, 1998] or apparently random restructuring of edges [Dash, 2003]. A typical example of reversal can be seen with a bicycle: In normal operation, you would apply force to the pedals which would cause the wheel to move, but you could instead manipulate the wheel directly and "release" the pedals, and then the causation of the system becomes reversed, the rotation of the pedals being caused by the wheel.

2.2 Difference-Based Causal Models

For some simple real physical systems, Dash [2003, 2005] showed that the Do operator would produce incorrect manipulation inferences. He traced these inconsistencies to the fact that when the equations underlying a strucural-equation-based causal graph are based on equilibrium relationships, then a manipulation on the system in question could take the system out of equilibrium and the old equilibrium relationship may no longer be valid. He called this a violation of Equilibration-Manipulation Commutability or the *EMC Condition*. The EMC condition is satisfied if a dynamic causal graph G that has been equilibrated ($\tilde{G}$) followed by being manipulated ($\hat{\tilde{G}}$) is the same as when those operations are reversed (i.e., when $\hat{\tilde{G}} = \tilde{\hat{G}}$).

EMC violation motivated a new representation, inspired by Iwasaki and Simon [1994], called *Difference-Based Causal Models* or *DBCMs* by Voortman et al. [2008]. DBCMs assume that all causation across time occurs due to a differential (difference) equation. DBCMs obey the EMC condition when the time-scale of the model is smaller than the time-scale of any process being modeled, i.e., small enough so that no equilibrations have occurred in a single time-step.

DBCMs differ from well-known temporal models such as Dynamic Bayesian Networks (DBNs) in that they force all causation to go through derivatives. Voortman et al. [2010] show that, when modeling very simple physical systems such as the simple harmonic oscillator, an algorithm that attempts to learn DBCMs as opposed to DBNs will produce infinitely more parsimonious models (i.e., the DBN model learned from data should have edges to the present from infinitely-many time-slices in the past), and they present a provably correct algorithm to learn them in the case of no latent confounding variables.

DBCMs are thus the only causal representation that we know of which can be guaranteed to obey the EMC condition, and thus are the "safest" models to use when we wish to reason about manipulation using the Do

operator. However, there are several problems associated with using them in practice. We discuss and make a first attempt at addressing these problems in the next section.

3 Toward an Improved Representation for Dynamic Models

When using DBCMs in practice, two key issues arise: the time-scale of the model, and the complexity of the model.

Getting a correct time-scale is critical to constructing a model that obeys the EMC condition, but how do we decide the right time-scale? Setting it as fine as possible would seem to be a safe choice. Ensuring it is fine enough so that no equilibrations have occurred may be a more practical choice. However, there are several practical difficulties with these strategies: First, it is both computationally prohibitive and intuitively unappealling to always model a macroscopic system at microscopic time-scales. It is much more convenient to select the operational time-scale that we are interested in (e.g., how frequently we wish to manipulate on the system), and let the causal model have both equilibrium and non-equilibrium relationships. Second, if we are interested in learning a causal system from data, we may not know the system well enough to judge what the correct time-scale is, and anyway we may have little if any control over the time-scale of the data we are using to learn our models. Third, there may be fundamental limits in our ability to gather fine-time-scale data and even to understand what the characteristic time-scale of a system sub-process is.

For example, consider the simple case of one body exerting a force on another. At macroscopic scales, Newton's first law dictates that the force F_{AB} on body A is equal and opposite to the force F_{BA} on body B. However, at microscopic time-scales, a much more complicated process is taking place: The crystal lattice of body A pushes against the lattice of B, causing tension in the chemical bonds of both lattices which feedback into the system. Eventually, $F_{AB} = F_{BA}$ is an equilibrium relationship that results. Trying to model this level of detail is inconvenient, computationally impractical, and may even run into fundamental limits set forth by Quantum Mechanics.

The second major issue with using DBCMs in practice has to do with model complexity: DBCMs, like DBNs, typically try to model a closed system, and therefore try to imagine all the relevant variables in the system *a priori*. They then produce a Markov model with repeating structure through time. While this produces a complete model that is capable of modeling manipulation with the Do operator, in a complex world where causes arise in response to unpredictable events, specifying a model that anticipates all possible causes, even in a closed but complex system, can quickly get intractable.

We attempt to address these two problems encountered for DBCMs. In order to do so, we seek a representation that is less sensitive to the observation time-scale, is sparser, provides better visualization of a specific instance of causation, and can be successfully manipulated. In order to accomplish all this, we return to the mechanistic view of causation.

Imagine a pool table with 16 balls on it with all balls sitting motionless for a duration of 5 minutes. We could imagine describing the causation of this system in at least two ways: One intuitive model is the simple Markov model stating that the position of the balls in time $t+1$ is simply given by the position of the balls at time t for all t in the time interval. Thus each ball's position in the future is caused only by it's position in the past. If at time $t = 5$ minutes ball A is struck with a cue for 0.5 ms,then we might describe the system by a temporal model, where at $t = 5$ minutes, an impulse force I arises in the system that causes A to accelerate, causing the velocity to increase, causing a friction force F to arise, eventually causing the ball to slow down and return to equilibrium. A sparse representation of this system is shown (for 2 balls) in Figure 1 (here we use the notation that $\dot{A}$ denotes the time-derivative of A).

A DBCM by contrast, would model this system in a more general way. Rather than treating the activity at $t = 5$ minutes as a special case, a DBCM would explicitely model the friction of each ball as a function of the velocity of the ball at all time slices. Then at early times, the velocity of the balls would be 0, and thus that causal path would have no effect on the dynamics. We might think of Figure 1 as a specific causal instance and think of the full-blown DBCM as a general model of the system. Taken that point further, when balls start colliding with each other and with walls, then at each time step, a DBCM model would require to prespecify all possible collisions that are possible. Doing this in a pairwise fashion may be feasible, but to imagine balls colliding in 3-way then 4-way collisions, and this model gets very difficult to maintain for even this relatively simple system. On the other hand, if we could develop a representation that took a more lazy approach to introducing causal relations, then we could produce a data-specific causal graph that was much sparser. The model in Figure 1 also does not demand that equilibrium relations be stricken from the model, as for the first 5 minutes the model consists exclusively of equilibrium relations.

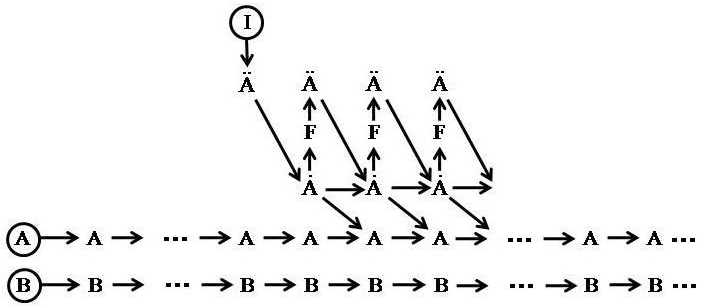

Figure 1:

As a description of a specific instance of a sequence of causal interactions, this new representations seems superior to DBCMs in that they are sparser, more informative (causal interactions can be deduced directly from the structure), and less fret with worries about getting the time-scale of the model correct. Two issues remain to be addressed for this representation: First, how do we use this "lazy" approach to model general (non-instance-based) causation? Second, how do we reason about manipulation?

First, general causation can be achieved with something akin to the mechanistic view of causation. Assuming we possess a library of mechanisms that operate on this system, and a set of rules for when those mechanisms get activated, then we could produce an algorithm for constructing a sequence of causal interactions from any set of initial conditions. For example, we may have a rule stating that when two balls collide, the conservation of momentum and mechanical energy relations must be added to the system. Adding these and restructuring the given time slice via the Causal Ordering Algorithm will allow us to produce the causal ordering for that time-slice. Manipulation can similarly be handled by the mechanistic view as the striking of some equations and replacing them with new equations, followed by a restructuring. Although we do not provide concrete answers to these questions here, we hope that listing the properties of the solution and revisiting the mechanism-based view of causality will lead to a new and better representation of causality.

References

Denver Dash. *Caveats for causal reasoning with equilibrium models*. PhD thesis, Intelligent Systems Program, University of Pittsburgh, Pittsbugh, PA, USA, April 2003. http://etd.library.pitt.edu/ETD/available/etd-05072003-102145/.

Denver Dash. Restructuring dynamic causal systems in equilibrium. In Robert G. Cowell and Zoubin Ghahramani, editors, *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AIStats 2005)*. Society for Artificial Intelligence and Statistics, 2005. (Available electronically at http://www.gatsby.ucl.ac.uk/aistats/).

David Hume. *A Treatise of Human Nature*. 1739.

Yumi Iwasaki and Herbert A. Simon. Causality and model abstraction. *Artificial Intelligence*, 67(1):143–194, May 1994.

Hans Van Leijen and Marek J. Druzdzel. Reversible causal mechanisms in bayesian networks. In *In Working notes of the AAAI 1998 Spring Symposium on Prospects for a Commonsense Theory of Causation*, pages 24–30, 1998.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2000.

Bertrand Russell. On the notion of cause. *Proceedings of the Aristotelian Society*, 13:1–26, 1913.

Herbert A. Simon. On the definition of causal relation. *The Journal of Philosophy*, 49(16):517–528, July 1952.

Steven Sloman. *Causal Models: How People Think about the World and Its Alternati ves*. Oxford University Press, USA, July 2005. ISBN 0195183118.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Springer Verlag, New York, 1993.

Robert H. Strotz and H.O.A. Wold. Recursive vs. nonrecursive systems: An attempt at synthesis; Part I of a triptych on causal chain systems. *Econometrica*, 28(2):417–427, April 1960.

Mark Voortman, Denver Dash, and Marek J. Druzdzel. Learning causal models that make correct manipulation predictions. *Journal of Machine Learning Research - Proceedings Track*, 6:257–266, 2008.

Mark Voortman, Denver Dash, and Marek J. Druzdzel. Learning why things change: The difference-based causality learner. In *Proceedings of the 26th Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*, 2010.

Herman Wold. Causality and econometrics. *Econometrica*, 22(2):162–177, April 1954.