Causal State Theory and Graphical Causal Models

Sarah Friedman[1]

[1]*2002 REU, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM, 87501*

(Dated: July 16, 2002)

An investigation of causal state theory and graphical causal models, as presented in:

1. "Computational Mechanics: Pattern and Prediction, Structure and Simplicity"
2. "Causality: Models, Reasoning and Inference" by Judea Pearl
3. "Causation, Prediction, and Search" by Sprites et al.

DIFFERENCES IN PHILOSOPHY AND GENERAL APPROACH

Computational Mechanics

What are patterns, and how should patterns be represented? Building from concepts and tools in statistical mechanics, causal state theory uses the data of a stationary time series with all joint probabilities known over blocks of all lengths. With this information, it is possible to reconstruct the ε-machine, a mathematical object embodying Occam's Razor in that its description of the process has minimal statistical complexity subject to the constraint of maximally accurate prediction. The ε-machine partitions the space of all histories of the process into sets of histories with the same conditional distribution of futures, i.e. causal states; it also provides the minimally-stochastic set of transition matrices, which give the probabilities of going from causal state i to j (while entering the symbols associated with that particular transition.) ε-machines are thus, in an abstract sense, the ideal representation of a process's structure, and the causal states are the minimal sufficient statistics for predicting the process's future. The problem is inferring ε-machines from limited data, and work of this approach now focuses on developing inference algorithms (state-merging versus state-creating) and accompanying statistical error theories, indicatiing how much data and computing time is required to attain a given level of confidence in an ε-machine with a finite number of causal states.

Graphical Causal Models

How might one extract causal relations from empirical, statistical data? How can (even incomplete) causal knowledge be used to influence and control systems around us (particularly social and economic systems)? This approach covers as much area as possible in the realm of mathematically expressing ideas of causal systems, understanding possibilities and limitations of determining causal structures from various kinds of data, and characterizing hypothesized efffects of a given intervention. Various types of graphs are defined, along with the conditions (such as Causal Markov, Minimality, Faithfulness, Stability) required for various degrees of niceness and predictive capability. The literature on graphical causal models is vast and underunified, despite attempts by different parties to give a definitive presentation of the theory. This approach also has had much work focusing on developing causal inference algorithms, for discovering causal relations between different events and displaying them graphically and nonrecursively.

WHAT IS CAUSALITY AND WHAT IS A CAUSAL RELATION? WHAT IS THE ALGEBRAIC OR REPRESENTATIONAL STRUCTURE USED TO ANALYZE OR DISPLAY CAUSAL INFORMATION?

Computational Mechanics

Causality is defined in temporal terms. Something causes something else by preceding it. In this way, the preceding state of affairs is considered to "produce" the present state of affairs. A chain of causal relations (a sequence of causal states, one leading to another) has the property of causal states at different times being independent, conditioning on the intermediate causal states. A causal state chain is Markov, in that knowing the whole history of the process up to a point has no more predictive capacity than knowing the causal state it's in at that point. The structure of causality is in the equivalence relation causal states establish among different histories which have the same conditional distribution of future observables, and in the probabalistic symbol-emitting transitions between the causal states. Directed graphs (vertices representing causal states) with labeled edges (represented emitted symbols and transition probabilities) can visually display the vital information in the ε-machine.

Graphical Causal Models

Pearl defines causality also in temporal terms (referrning to Hume, etc.), along with an emphasis on inter-

vention as the true test of a hypothesized causal relation. Sprites et al refuse to come out and define causation or causality, instead outlining three views of the nature of causation: 1. Causal influence is a probabalistic relation. 2. Causal influence is a counterfactual relation (ideas of manipulation or intervention - Pearl may fall under this view). 3. It is better not to talk of causation. They then describe causation as being transitive (A causes B and B causes C implies A causes C), irreflexive (A connot cause itself) and antisymmetric (if A causes B then B cannot cause A). I find the second two properties to be quite unfortunate in terms of failing to describe many causal relations of scientific interest, but they represent the theoretical basis for the graphical approach's representational structure. The formalism most commonly used to express causal relations is that of directed graphical models (vertices to denote events - often defined by Boolean variables - and directional edges to denote cause-and-effect relations). The absence of causation can generally be recognized by independent probabilities - the many graphical-approach causal inference algorithms generally take these ideas as their conceptual basis. Because of the problems of discovering mutual causation from empirical statistical data, as well as the very definition of causation in the case of Sprites et al., all causal models are represented as DAGs. A justification lies in Basmann's (1965, cited in Sprites et al.) argument that for every simultaneous equation model with a cyclic graph, there exists a statisitcally indistinguishable model with a acyclic graph; also, as previously stated, acyclic graphs are decidedly less problematic to infer statistically with present tools such a Bayesian analysis.

WHAT IS THE FORM OF THE DATA USED FOR INFERRING CAUSALITY? EQUIVALENTLY, WHAT PROCESSES ARE BEST CAPTURED BY THE METHOD?

Computational Mechanics

Time series or other sequential data, in the form of a stream of symbolic outputs. The ε-machines depend on the existence of some alphabet of symbols from which is taken the symbol emitted at each step. Also, the process must be stationary, i.e. in some sense the causal states are not changing over time. (This condition could conceivably be relaxed, especially when using an "on-line" algorithm.) Transient states are trimmed off (although they can be easily rediscovered) in the search for recurrent states, to which the bi-infinite process would return infintely often. The longer the time series or symbol sequence data, the smaller the alphabet, and the longer the string of symbols the algorithm can consider as a suffix at the end of a history the more accurate the ε-machine reconstruction. Obviously, one assumes the process has a finite number of causal states. The processes best captured by this method are sequences of a symbolic logic character, with mysterious and complex pattern and an unknown generative mechanism, such as an extremely long string on not-totally-random 1's and 0's that are an output of the golden mean system.

Graphical Causal Models

Although one of the main ideas of this approach is to be able to deal with as many forms of data as possible, the most general form of data seems to be a collection of statistics on different potential causal factors and as much data as possible on the associated joint and conditional probabilities. Generally the data is not sequential or in a time series, but quite often one can use common sense to determine some kind of partial sequential ordering (i.e. a wet sidewalk cannot cause rain, but rain causes a wet sidewalk). Importantly, when different records or distributions are mixed, there can be a spurious vanishing of association in the statistics (Simpson's "paradox"), or when exogenous causal variables are not accounted for, the causal graph may make no sense. One must be sure not to mix records or omit relevant latent (unobserved or unobservable) variables. Also, just because some equations of functional dependencies correctly describe a system does not mean that one can infer direct causal dependencies from those equations. (However, knowing causal dependencies does allow one to infer functional dependency equations.) There are usually many different graphs that accurately describe the available data, so the graphical approach outlines quite a lot of different definitions of types of graphs, and conditions for which graphs are preferable when, and why. The field seems pretty incoherent to me on that point, but the work seems to boil down to the search for an ideal causal inference algorithm that can deal with many different types of data and systems and give the most certain graph with the lowest probability of error. The processes best captured by this method are somewhat large-scale socioeconomic phenomena, such as the relative effects of sex, IQ, parental encouragement and socioeconomic status on college plans.

(Note for what follows: "causality" and "causation" are synonyms according to the American Heritage Dictionary.)

HOW TO REPRESENT RANDOMNESS?

Computational Mechanics

Randomness is orthogonal to structure. Causal states describe the structure of a process with maximum prescience and minimal statistical complexity. Randomness

is embodied in the minimally stochastic transitions form one state to another. (The given state can lead to many states, but a given symbol leads to at most one of those states.)

Causal Graphical Models

It depends on the form of the model. Bayesian networks (which are stochastic) seem to just graphically display information about joint probabilities, perhaps making them something of a simplistic ε-machine without symbol emissions from transitions. The deterministic and pseudodeterministic causal models account for randomness in the form of a mystery node denoted u (for "unobservable") which affects the outcome of a transition from one observable node to another.

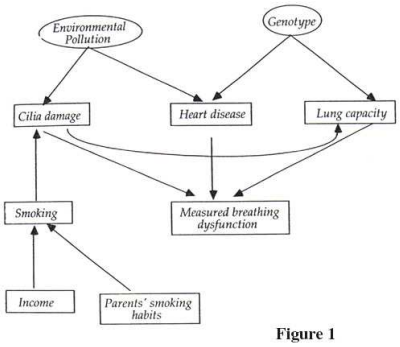

FIG. 1: Causal Structure of Breathing Dysfunction

EXAMPLES TO DISPLAY CONTRASTING FOCI AND STRENGTHS OF BOTH APPROACHES

The Even Process (from a draft of "An Algorithm for Pattern Discovery in Time Series Part 1")

This is a process which generates a string of 1's and 0's according to the rule that after a 0 is emitted, there is a coin flip to see whether a 1 or 0 follows. After an odd number of 1's being emitted, the process must emit a 1, but after an even number of 1's it is back to the coin flip. ε-machine reconstruction (with a causal-state-splitting reconstruction) uses the data of a sequence of 10^4 1's and 0's that is the output of one run of the even process, and comes up with the correct causal states, and transition probabilities that are within .3% of the correct probabilities. It is difficult to determine what the graphical model approach would do, since there are so many different possibilities. A fundamental problem is how to define the "events" in question, since the data is not in terms of potential factors. I suppose the natural thing would be to say there are two events, either "emit a 1" or "emit a 0," and then calculate the probability of "emit a 0." You'd look at "emit a 1" given "emit a 1" and "emit a 0," and you might not even be able to represent the sequential nature (that "emit a 1" followed "emit a 0") because of the way these algorithms use Bayesian analysis and don't recognize time series. This is because of the requirement to represent the process as a DAG. Bayesian analysis to reconstruct the DAG would lead to an infinite lattice of similar events and probabilities. Hopefully at some point the modeler would have the sense to see what was going on, and perhaps choose to break with the herd and represent the process as a DAG - the process is simple enough that it might be possible to see what was going on. It would require human thought and effort to do that, because none of the algorithms would be able to recognize the cyclic temporal causal relations.

Causes of Breathing Dysfunction (from Sprites et al., ppl 130-132)

This is an imaginary but somewhat realistic example. See fig. 1 for the "real" causal structure, from which the marginal distribution over the boxed (endogenous) variables was generated. Environmental Pollution and Genotype are exogenous common causes of Cilia damage, Heart disease, and Lung capacity, which are all contributing causes of Measured breathing dysfunction, and Cilia damage is also caused by Smoking which is caused by Income and Parents' smoking habits.

ε-machine reconstruction would not be able to deal with this data. The data could perhaps be put in the form of a time series of symbols describing the condition of a subject's breathing capacity. It could then reconstruct causal states and transitions, but the only information I can imagine being captured would be that breathing capacity tends to get worse and worse through time, perhaps reflecting the cumulative effects of smoking and environmental pollution. The fact that only one subject was being used would preclude any capture of individual-specific causal factors such as parents' smoking habits or genotype. All in all, ε-machines could not uncover the underlying causal structure of the process, being unable to account for these variable causal factors.

Sprites et al. cover two different algorithms' reconstruction of the causal structure (see fig. 2 and fig. 3). One thing I don't understand is why there are a couple two-headed arrows - these are the guys who defined causation as being antisymmetric, but such quirks are typical of this approach. The circles on some edges denote unexplained correlation. The Fast Causal Inference algorithm seems to do a better job of uncovering the causal structure, but both algorithms do a relatively good job. Although there is a lot of uncertainty, there are no mistakes.

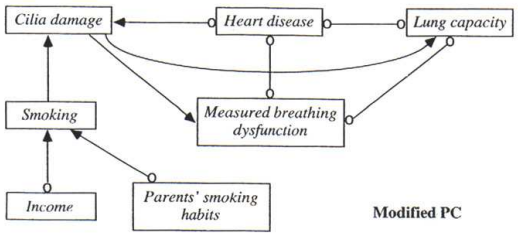

Figure 2

FIG. 2: Reconstructed Causal Structure of Breathing Dysfunction 1: Modified PC Algorithm

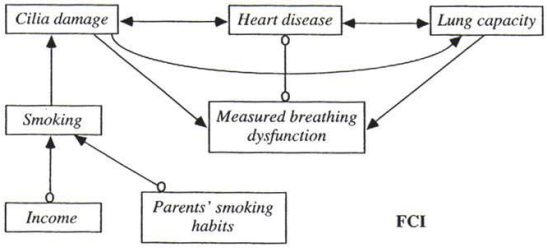

Figure 3

FIG. 3: Reconstructed Causal Structure of Breathing Dysfunction 2: Fast Causal Inference Alogorithm

THE CONTRADICTIONS BETWEEN AND LIMITATIONS OF BOTH APPROACHES

Computational Mechanics

The philosophy of quantifying Occam's Razor could be seen as a closed world assumption in Pearl's view, if you think it is implicitly assumed that all relevant factors have been accounted for. On its own terms, however, the causal state definition sidesteps this, because it doesn't matter what determines the causal state to be a causal state, it just depends on the future morph of a given history. "Factors" is not a meaningful term for the definition of an ε-machine. Also Pearl's notion of "stability" - that the model's structure can retain its independence pattern in the face of changing parameters - (a kind of robustness) is not satisfied by the ε-machine reconstruction, because it is unclear what the "parameters" are. In effect, the causal state theory can show the likelihood of a certain symbol being emitted, given a history (or a causal state), after reconstructing the ε-machine from the exact joint probabilities over sequence blocks of all lengths (data from *one run*). However, it doesn't display causal factors, and genuine causation versus potential causation versus spurious association, the way graphical causal models do. In this way causal state theory seems comparatively impractical for use in fields where qualitative causal assumptions have value, where causal factors are vitally important for intervention (policy purposes, etc.) yet all relevant factors cannot be accounted for, and/or data is in the form of a record of statistics on related potential factors rather than as a time series. (This is also just a matter of approach. Graphical models are trying to extract causal relations from data in order to predict effects of various possible interventions, CM is trying to understand causal patterns in the abstract and represent them formally and rigorously - both do an impressive job, but they have different fundamental tasks which leads to these contradictions and limitations.)

Graphical Causal Models

Both graphical books' model conceptions rely completely on DAG's as the underlying structure. Directed Acyclic Graphs are similar to causal state graphs, having marked and directed links. However, a major shortcoming of DAG's in terms of representing actual processes is their innate inability to represent recursion, mutual causation or feedback loops (by being "acyclic"). This problem does not come up for graphical causal models to have universal use-value, since auto-catalytic and homeostatic processes are pervasive and embedded in so many processes of scientific interest. Also, in its pursuit of full coverage and generality the graphical approach seems to sacrifice coherence and clarity of theory. The problems tackled by this approach live in the extremely complex world of health, social, and economic systems, where it is extremely difficult to find large, reliable datasets, let alone account for all factors, leading to intrinsic analytical issues in the field.

POSSIBLE CONNECTIONS

Pearl regards the idea of "intervention" as fundamental to causal inquiry. Some of his definitions, such as his definition of Causal Bayesian networks, relay on the operation do(X=x) to make causal statements. With the definition of a causal state as all histories having the same morph, intervention does not seem as important. An intervention could be represented by setting an appropriate

$$\mathrm{T}_{ij}^{(s)} = 1 \ . \tag{1}$$

Also, the interventions seem to serve the purpose of discerning conditional independencies, which the ε-machine reconstruction takes as given. So probably when you're

at the point of reconstructing the ε-machine, you've done all the interventions you need to do to determine joint probabilities.

A nice connection is that DAG nodes are independent, conditioning on the intermediate nodes; just as causal states are independent, conditioning on the intermediate causal states. This notion seems to be universal to all concepts of causal chains. (Again, since DAG's can't capture mutual causation or recursion, ε-machines seem to be more general.)

Another idea that both approaches share (for obvious reasons) is the focus on Minimality (or minimal statistical complexity) as a main goal of a causal model. Nobody wants to "overfit" the data.

For Pearl, *functional* causal models are deterministic (like ε-machines), and are represented by a set of equations of the form

$$x_i = f_i(pa_i u_i) \tag{2}$$

where pa denotes "parents" and u denotes "unobserved disturbances," so x_i is a random variable determining an observed or observable value. If each equation represents an autonomous mechanism that determines the value of just one distinct variable, then the model is a structural causal model a.k.a. *causal model*. If we take "parents" to represent "prior causal state" - and "unobserved disturbances" and f to somehow account for or determine the transition between the prior and present causal states - then we could characterize any causal model as a (perhaps a very rough and definitely nonrecursive) ε-machine, or equivalently, an ε-machine could be constructed out of a causal model.

BIG PROBLEMS

ε-machines do not try to help up figure out causal problems of major public interest, such as what we could do to decrease cancer rates. Graphical models have major analytical weaknesses (incoherent theory, inability to deal with causal symmetry and reflexivity). I'm not sure if it's possible in the present circumstances to combine these two approaches in a Unified Causal Theory that lets us tackle many kinds of relevant policy issues with analytical rigor and clarity, but it seems that the common goal should be something of that nature.