Learning Across Space, Time, and Scale: A Bayesian Perspective

M. Shane Tutwiler, Tina A. Grotzer, Harvard University, Cambridge MA
Email: mst216@mail.harvard.edu, tina_grotzer@harvard.edu

Abstract: Theories of causal inference and pattern recognition based on machine learning have been proposed as normative models of human learning. To date, these theories fail to include explanations for why humans are biased towards some types of data (such as surprising or confirming) over others. In this poster we will provide a novel explanation for this, and use this hybrid theory to highlight areas of prior CSCL research that successfully supported student learning across space, time, and scale, as well as propose future research.

Introduction

Humans of all ages frequently make rapid causal inferences based on covariational information. In simple settings, when effects and probable causes are well defined, this process is relatively efficient. For example, assume that you own a cat with a particularly weak constitution. One day, said cat eats some food that you drop on the ground and later becomes ill. Does that cat have a virus, or do you presume that its current condition is the result of its having eaten some of your lunch? Normative theories of causal inference predict that you will use relative prior information about your cat (such as how frequently it gets ill after eating your food vice how frequently it is sick with a viral information) to make a decision about taking it to the vet. In this poster, we will give a brief overview of these theories of human causal inference, and then combine them in a novel way with theories of information processing from computer science in order to propose best practices and future strands of CSCL research regarding causal learning across space, time, and scale.

Background

Causal Bayes Nets

Developmental psychologists and cognitive scientists have long endeavored to quantify and predict human causal inference. Early models were based on relative frequencies (i.e. covariation) of candidate events and causes (Jenkins & Ward, 1965). Eventually, these equations were modified to encompass measures of causal strength as well (Cheng, 1997). These models were limited in the assumptions they made about cause-effect relationships, however. Glymour (1998) re-framed Cheng's (1997) theory as a type of graphical analysis known as a Causal Bayes Net (CBN). Based on the work of Pearl (1988) the CBN theory of human causal inference posits that, given well defined inputs, humans will make simple causal inferences in a normative manner. We can use the scenario in the introduction, above, as an example (Figure 1, below).

$$V \xrightarrow[\omega 0]{} S \xleftarrow[\omega 1]{} F$$

Figure 1. Causal Bayes Net depicting the causal effect of a virus (V) and/or food (F) on your cat being sick (S) (ω0 and ω1 represent the strength of the observed causal connections.)

Observing Figure 1, above, we note that a directed edge (arrow) goes from V to S and from F to S, indicating that both V and F are independent causes of S. In addition, we note that each causal relationship has an associated weight (ω) that indicates the strength of the proposed connection, based on observation of covariation. In the scenario above, you would be likely to "screen off" V as a potential cause if the strength of the relationship between F and S was historically stronger. This highlights two points: 1) humans may infer simple causal relationships in a manner consistent with CBNs, and 2) this inferential process may prohibit humans from identifying complex causal relationships (Grotzer & Tutwiler, in preparation).

Researchers over the last decade have used CBN-based theories to model and predict human causal inference across different tasks and ages (e.g. Schulz & Gopnick, 2004; Griffiths &Tenenbaum, 2005). These studies all focus on learning simple relationships, however, and for good reason. The developers of CBN methods (Pearl, 1988, 2000), have shown that computation of CBNs become intractable as systems become more complex (Bishop, 2006). If the core cognitive mechanism used to infer causal connections may force people to form overly simplistic causal models, and that the mechanism can't be scaled up directly to infer more complex models, how does this theoretical framework help us to positively impact student causal (or systems) learning?

Information Theory

One way to help support student understanding of complex systems, assuming a CBN paradigm, is through the application of information theory. The degree of surprise in learning some piece of data "x" is defined as *information*, and is given in column 1 of Table 1, below. In essence, the higher the information level some piece of data has, the more likely one is to use that data in the causal weights (ω) defined above. The *amount* of information you have to transmit in order for "x" to be used in the updating of your update beliefs is said to be the *entropy*, and is given in column two of Table 1, below. In general, entropy increases as distributions become more broad and uninformative. In other words, if students are exposed to data that they already expect (*or that they don't know not to expect*), then it takes *much* more data to override a prior belief. Finally, we also consider the relative entropy (Table 1, Column 3), or the average addition information required to transmit a value (x) if we assume a distribution, q(x), which is not exactly the same as the distribution actually generating the data, p(x). In other words, if the student has the wrong model in mind, the amount of extra information, on average, that they have to gather before they correctly discern the value of x, which can then be used to update their prior belief, is the *relative entropy*. In essence, the *amount of information needed becomes much greater as p(x) and q(x) diverge.*

Table 1. Data information, entropy, and relative entropy equations

Information	Entropy	Relative Entropy
$h(x) = -\ln p(x)$	$H[x] = -\sum_{x} p(x) \ln p(x)$	$KL(p\|\|q) = -\int p(x) \ln\{\frac{q(x)}{p(x)}\}dx$

Bayesian Updating

The updating of causal weights with new data described above can be framed in terms of belief change using Bayes' Theorem: $P(\omega|D) = P(D|\omega)P(\omega)/P(D)$

In effect, your belief in the strength of a relationship, given new data, is proportional to your prior belief in that relationship and the likelihood of that new data. If the data is high information (surprising, or low entropy) or if your mental model closely aligns with the process that is generating the data, then you will be more likely to update your belief, which will then impact the model (CBN) you use to evaluate future data.

A CBN/IT Perspective of CSCL

Assuming that students make causal inferences in ways consistent with CBN theories, weigh information in ways consistent with Information Theory, and generally update prior beliefs in a Bayesian manner, then certain best practices can be recommended. In our poster, we highlight examples from prior CSCL research in which data Information, Entropy, and Relative Entropy were properly leveraged to maximize causal (or systems) learning across instances of space, time, and scale. These examples include studies of a multi-user virtual environment (Metcalf et al, 2011), hypermedia (Liu & Hmelo-Silver, 2009), and role-play (Deaton & Cook, 2012). The theoretical contributions from this poster should help to inform future research on causal and systems learning research, and should be of great benefit to the CSCL community.

References

Bishop, C.M., (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer Science+Business Media, LLC.

Cheng, P.W. (1997). *From covariation to causation: A causal power theory. Psychological Review, 104,* 367-405.

Deaton, C.C.M. & Cook, M. (2012). Using role-play and case study to promote student research on environmental science. Science Activities 49(3): 71-76.

Glymour, C. (1998). Learning Causes: Psychological Explanations of Causal Explanation. *Minds and Machines*, 8(1998), 39-60.

Griffiths, T.L. & Tenenbaum, J.B. (2005). Structure and Strength in Causal Induction. *Cognitive Psychology, 51, 334-384.*

Grotzer, T.A., & Tutwiler, M.S. (in preparation) Causal Bayes Nets: A bridge too far?

Liu, L. & Hmelo-Silver, C. E.(2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. Journal of Research in Science Teaching, 46, 1023-1040 .

Metcalf, S.J., Kamarainen, A., Tutwiler M.S., Grotzer, T.A. & Dede, C. J. (2011). Ecosystem science learning via multi-user virtual environments. International Journal of Gaming and Computer-Mediated Simulations. 3(1)86-90.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* San Francisco, CA: Morgan Kaufman Publishers, INC.