The Temporal Logic of Token Causes

Samantha Kleinberg and **Bud Mishra**
New York University
715 Broadway, 10th floor
New York, NY, 10003

Abstract

While type causality helps us to understand general relationships such as the etiology of a disease (smoking causing lung cancer), token causality aims to explain causal connections in specific instantiated events, such as the diagnosis of a patient (Ravi's developing lung cancer after a 20-year smoking habit). Understanding why something happened, as in these examples, is central to reasoning in such diverse cases as the diagnosis of patients, understanding why the US financial market collapsed in 2007 and finding a causal explanation for Obama's victory over Clinton in the US primary. However, despite centuries of work in philosophy and decades of research in computer science, the problem of how to rigorously formalize token causality and how to automate such reasoning has remained unsolved. In this paper, we show how to use type-level causal relationships, represented as temporal logic formulas, together with philosophical principles, to reason about these token-level cases.

Introduction

When we want to determine what is responsible for a patient's symptoms, why a stock plummeted in value, or the reason a particular candidate won an election, what we want to know is what *caused* these particular events. But rather than finding a general relationship, such as "smoking causes lung cancer", we want to find whether a particular one, such as "Bob's smoking caused his lung cancer" is true. In order to do this in an automated way, we need to understand the general relationships (called type-level causality) and how these relate to the singular cases (called token-causality). We also need a system for combining this knowledge in a rigorous, automated, way.

While the problem of general causal inference has been studied in our prior work (Kleinberg and Mishra 2009) as well as that of other computer scientists, we cannot immediately use these inferences to explain token cases. A type-level relationship may indicate that a token case is likely to have a particular cause, but it does not necessitate this. Just as the relationship between smoking and lung cancer does not mean that all lung cancers are caused by smoking, we cannot immediately propose that a type-level cause is a token-cause. We must first establish whether the type-level relationship has been instantiated and take into account that we may wish to also assess the role of other hypotheses, including rare factors. In some cases we will not even know if the potential cause occurred, we may only have indirect information such as whether its causes and effects occurred. We discuss the general problem and approach here, with more details and examples provided in the full paper (Kleinberg and Mishra 2010).

Computational approaches have traditionally looked at the problem of beginning with a type-level model (such as a Bayesian network), and then using this to assess a particular case. Approaches in logic have focused on the problem of reasoning about the results of actions on the system (Lin 1995; Thielscher 1997) or diagnosing the causes of system malfunctions based on visible errors (Lunze and Schiller 1999). Most recently, Hopkins and Pearl (2007) have proposed a framework drawing on earlier work on structural models as well as the work on situation calculus. In this more recent adaptation, it is shown that counterfactuals may be modeled using the situation calculus, however one must still specify all dependencies - including those of counterfactuals.

The relationship between type and token

Previous approaches require that one must either begin with a model, know the truth values of all variables, or have a deterministic system. In contrast, we will infer relationships (temporal logic formulas with a causal interpretation) from time series data and then assess the support of each of these hypotheses for a token case.

Type-level inference

We will give a brief overview of our approach to type-level inference before discussing how to use these type-level causes for token-level cases. In prior work (Kleinberg and Mishra 2009) we created a new framework for causal inference, where cause and effect are described in terms of probabilistic computation tree logic (PCTL) formulas (Hansson and Jonsson 1994), and checked to see if they are satisfied in time series data (traces) using model checking. Then, to determine which of these possible inferred causal relations are significant, we compute the average difference a cause makes to its effect, using the concept of multiple hypothesis

testing to determine at what level something is statistically significant (Efron 2004).

Definition 1. A factor c is a *prima facie* cause of e if $P(c) > 0$ and the probability of e, when c occurs prior to e is greater than $P(e)$.

In order to determine whether a particular *prima facie* cause is significant, we compute, with X being the set of all other *prima facie* causes of e:

$$\epsilon_{avg}(c, e) = \frac{\sum_{x \in X \setminus c} \epsilon_x(c, e)}{|X \setminus c|}, \quad (1)$$

where

$$\epsilon_x(c, e) = P(e|c \wedge x) - P(e|\neg c \wedge x). \quad (2)$$

Then we use this ϵ_{avg} to determine c's significance.

Definition 2. A prima facie cause, c, of an effect, e, is an *ϵ-insignificant cause* of e if $\epsilon_{avg}(c, e) < \epsilon$.

Definition 3. A prima facie cause, c, of an effect, e, that is not an *ϵ-insignificant cause* of e is an *ϵ-significant*, or *just-so*, cause.

The connecting principle

We will now use the strength associated with the type-level causes to assess the strength of token-level claims using the *connecting principle*, introduced by Sober (1986). The basic idea is that the support of a particular token hypothesis (such as Bob's smoking caused his lung cancer) is proportional to the strength of the type level relation (such as smoking causes lung cancer). That is, if we know of a type-level relationship between C and E, then the magnitude of the support for C token causing E, given that the both actually occurred, is proportional to a measure of strength for the related type-level relationship. The main principle here is that a known type-level relationship between some c and e is good evidence for c causing e, if we see that both c and e have occurred.

Token-level reasoning

We will now reframe Sober's principle for our purposes, using our measure of significance and allowing incomplete information.

The set of possible token causes

Recall that we are calculating the support of token causal claims - with the assumption that we are interested in those with high levels of support. If two possible token causes took place on a particular occasion and one is significant while the other is insignificant, the more likely explanation for the effect is that it was token caused by the type-level significant cause. That is, if we have a number of token causal hypotheses, those with the highest support will be those with the highest value for ϵ_{avg} – our just-so or significant causes. Thus, if we know that a just-so cause of the effect in question took place, we do not need to examine any insignificant or non-prima facie causes of the effect, as the only other causes that may have higher significance for the effect are other significant ones. If none of the just-so or significant causes occurred, then at that point we would have to examine alternative hypotheses.

Support of a causal hypothesis

We will not always know whether a cause occurred, so we now reformulate Sober's measure of support to account for this.

Definition 4. Assume that e token-occurred in population P; that the probability that c token-occurred in P is $P(c)$; and that $\epsilon_{avg}(c, e)$ is the strength of the type-level relationship between c and e. Then, the support for the hypothesis that c token-caused e in P is:

$$S(c \rightsquigarrow e) = \epsilon_{avg}(c, e) \times P(c). \quad (3)$$

This means that if we have full knowledge of a scenario, the support for each possible explanation will be exactly equal to the strength of the corresponding type-level relationship. However, when we have missing data and are unsure as to whether or not a possible cause occurred, the support for the hypothesis will be weighted by the probability of the cause having occurred, given what we have observed.

Calculating the probability of c

To calculate the probability of a particular cause token-occurring, we can go back to our original data, using frequencies (calculating the frequency of sequences of length t where the evidence holds). However, if we have or have inferred the structure of the system, we may use that instead. In either case, note that we are computing the posterior probability of c, where our evidence is one sequence of observations, comprised of a conjunction of the facts about the scenario. We will refer to this evidence as E. We are interested in $P(c|E)$, which is by definition:

$$P(c|E) = 1 - \frac{P(\neg c \wedge E)}{P(E)}. \quad (4)$$

The facts we have about the current scenario will be time-indexed such that we have facts at times t_1, t_2 and so on, indexed relative to the beginning of the event or at times such that we know their order and can calculate the elapsed time between them. These facts constrain the set of states our system has occupied (assuming our model of the system is correct, or our data is representative of the system). We may also limit the evidence considered to known causes and effects of c. Note that there are time windows associated with all the causal relationships, and thus when we say $\neg c$, we mean c did not occur in such as way as to satisfy the formula representing its relationship with some e (i.e. c did not occur within the correct time window).

Thus, when we do not have a model, the probability, $\frac{P(\neg c \wedge E)}{P(E)}$, will be the number of times the sequence of facts in F' is true along the trace, divided by the number of times the sequence F is true. For a set of traces, these would correspond to the number of traces in which each set holds. When we have a model, we calculate the probability of this

sequence of conjunctions holding. We will repeat this procedure twice, once for the numerator and once for the denominator, thus calculating the probability of each cause having occurred using Equation (4).

Procedure for assigning support to causes

Recall that we have sets of type-level causes of the token-effect in question. In order to determine the support for each, we must first ascertain - using the facts about the situation - which of these occurred. When we do not have enough information to determine if one has occurred, we use the above procedure to determine its probability using our observed evidence. Recall that the support for each hypothesis is the previously computed ϵ_{avg} - weighted by the probability of the evidence. That is, the largest possible value of the support for a token hypothesis is its associated ϵ_{avg} (since the probability can be at most one). If any significant type-level causes have occurred, this means that they will have the highest values of this support. With C being the set of significant causes of the token-effect, e, and F being the set of known time-indexed facts, we test whether each $c \in C$ is true on this occasion given the facts. This means determining whether the components of the formulas occurred in such a way as to satisfy the causal relationship. Thus if the formula is $q \leadsto^{\geq 1, \leq 2} e$ and we know q at t_1 and e at t_2, the formula would be true in this token instance, while if the facts were instead that q at t_1 and e at t_4, it would be false. When the support for the significant hypotheses is very low or zero, we must examine the other potential explanations: our previously discarded insignificant causes, and perhaps those that are not even prima facie causes. We may define a threshold at which we will examine less likely causes. The result is a set of possible explanations ranked by their support, with those having the highest values being the preferred explanations for the effect. We can also test any hypotheses of interest to see how they relate to the token effect.

Example

We now discuss a simple example, illustrating the approach. More examples, including difficult cases, appear in the full paper (Kleinberg and Mishra 2010). We begin with Bob and Susie, who are each armed with rocks that they may throw at a glass bottle. Let us assume we have already found one type level significant causes (with all other causes being insignificant) of such a bottle breaking in this system. This is represented by:

$$T \leadsto^{\geq 1, \leq 2}_{\geq p} G. \quad (5)$$

That is, throwing (T) a rock from a certain distance causes the glass to break (G) in greater than or equal to one time unit, but less than or equal to two time units, with at least probability p. Since we have found this to be a type-level cause, we have the associated values of ϵ_{avg}.

We have the following facts about the token case:

1. Bob threw his rock at time 3;
2. Susie threw her rock at time 4;
3. The glass broke at time 4;
4. The only significant cause of a broken glass is that in formula 5.

Our type level relationship says that if T is true at some time t then it can lead to G being true at time $t + 1$ or $t + 2$. The facts we begin with are that Bob's instance of T is true at $t = 3$ and Susie's at $t = 4$. To satisfy the causal formula of (5), G would need to be true at $t = 4$ or $t = 5$. G is true at 4 and thus Bob's throw can be considered as a possible token-cause of G. Now, for Susie's throw to be a token cause of G, G would need to be true at $t = 5$ or $t = 6$. However, G is true at $t = 4$, which means this causal relationship did not occur, and it is not a possible token cause (since it could not lead to G at the time at which G actually occurred). Thus in this case our only potential token cause is Bob's throw, and the support for this token cause will be $\epsilon_{avg}(T, G)$.

Conclusion

We have shown how inferred type-level causes, represented by logical formulas, may be used to reason about token-level cases. This method captures information about the timing of the general relationship and occurrence of actual events, allowing automated reasoning about cases that were previously only correctly handled with intuition. In future work we will discuss how to include other knowledge as well as the possibility that some "facts" may be conflicting or incorrect.

References

Efron, B. 2004. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association* 99(465):96–105.

Hansson, H., and Jonsson, B. 1994. A logic for reasoning about time and reliability. *Formal Aspects of Computing* 6(5):512–535.

Hopkins, M., and Pearl, J. 2007. Causality and Counterfactuals in the Situation Calculus. *Journal of Logic and Computation* 17(5):939.

Kleinberg, S., and Mishra, B. 2009. The Temporal Logic of Causal Structures. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Kleinberg, S., and Mishra, B. 2010. The Temporal Logic of Token Causes. Technical Report TR2010-926, CIMS New York University.

Lin, F. 1995. Embracing causality in specifying the indirect effects of actions. In Mellish, C., ed., *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1985–1991. San Francisco: Morgan Kaufmann.

Lunze, J., and Schiller, F. 1999. An example of fault diagnosis by means of probabilistic logic reasoning. *Control Engineering Practice* 7(2):271–278.

Sober, E., and Papineau, D. 1986. Causal factors, causal inference, causal explanation. *Proceedings of the Aristotelian Society, Supplementary Volumes* 60:97–136.

Thielscher, M. 1997. Ramification and causality. *Artificial Intelligence* 89(1-2):317–364.