A Circumscriptive Theory for Causal and Evidential Support

Eunok Paek
Department of Computer Science
Stanford University
Stanford, California 94305
paek@cs.stanford.edu

Abstract

Reasoning about causality is an interesting application area of formal nonmonotonic theories. Here we focus our attention on a certain aspect of causal reasoning, namely *causal asymmetry*. In order to provide a qualitative account of causal asymmetry, we present a justification-based approach that uses circumscription to obtain the minimality of causes. We define the notion of causal and evidential support in terms of a justification change with respect to a circumscriptive theory and show how the definition provides desirable interactions between causal and evidential support.

1 Introduction

Commonsense reasoning has provided a rich application area for nonmonotonic logic. As some researchers have noted, however, certain aspects of causal reasoning which are prevalent in ordinary discourse have not received due treatment by nonmonotonic logic[Pearl, 1988a].

In this paper, we will focus on the phenomenon of *causal asymmetry* which results because two causes of an observation interact differently than two consequences of a common cause. In [Pearl, 1988a], Pearl presents a causal reasoning system as an attempt to provide a qualitative account of such causal directionality shown in his Bayesian analysis[Pearl, 1988b]. His system, however, generates certain counterintuitive conclusions, as Section 3 will illustrate.

Here we introduce a better qualitative account of causal directionality that overcomes the difficulty mentioned in Section 3. Our approach is not as detailed as the probabilistic account of causal directionality, but a qualitative approach has its advantages: it is simpler and more intuitive than the probabilistic approach.

In Section 2, we will briefly describe Pearl's causal reasoning system and clarify a few implicit assumptions made by the system. In the following section, we will look into some difficulty with his framework. In Section 4, we present a simple circumscriptive theory[McCarthy, 1986] that allows us to draw explanatory conclusions. Finally we will give a justification-based account of causal directionality using this circumscriptive theory and also discuss how our approach improves on Pearl's.

2 Pearl's Causal Reasoning System

In [Pearl, 1988a], Pearl states that it is necessary to know whether a given proposition was established by causal or evidential considerations, and that there is a causal asymmetry stemming from the fact that two causes of an observation interact differently than two consequences of a common cause; in the absence of a direct relation between the two, the former *compete* with each other, while the latter *support* each other. For example, when we observe a rash we are more likely to observe fever as well because measles always involves both fever and rash. In contrast, when we know that the cause of a rash is measles, it is less likely that the patient is also suffering from an allergy. In order to produce such an asymmetry, Pearl proposes a reasoning system in which every proposition is labeled as either *causally* or *evidentially established*, every rule is labeled as either *causal* or *evidential*, and inference rules are defined according to these labels. That is, each rule in the system is labeled as C (connoting "causal") or E (connoting "evidential"), and $P \rightarrow_C Q$ means "P causes Q" and $Q \rightarrow_E P$ means "Q is evidence for P". Similarly each proposition is labeled either as E or C where $E(P)$ means that P is believed for evidential reasons and $C(Q)$ means that Q is believed for causal reasons. The semantics of the $C - E$ distinction are defined by the following three inference rules

$$\frac{P \rightarrow_C Q \quad C(P)}{C(Q)} \qquad \frac{P \rightarrow_C Q \quad E(P)}{C(Q)} \qquad \frac{P \rightarrow_E Q \quad E(P)}{E(Q)}$$

while purposely precluding the inference rule

$$\frac{P \rightarrow_E Q \quad C(P)}{Q}$$

Before we explain how these inference rules deal with causal asymmetry, we want to clarify a few assumptions made by Pearl's reasoning system. First, in his reasoning system it is not specified how we initially obtain propositional labels. When we have some labels for certain propositions initially, we can deduce other labeled propositions by using the initially labeled propositions and rules according to the inference rules sanctioned by the system. Without any initial labeling, however, we cannot use the inference rules at all. But from the fact that we can use a causal rule $P \rightarrow_C Q$ together with a proposition P regardless of its label, we can obtain labels for certain propositions without any initial labeling for P. From now on, whenever we have a rule $P \rightarrow_C Q$ and a proposition P without any label, we will assume that $C(Q)$ follows. Secondly, there is an implicit correspondence between causal and evidential rules. That is, the database must have $Q \rightarrow_E P$ whenever it has $P \rightarrow_C Q$. If we can represent evidential rules without assuming the existence of their causal counterparts, some inference rules can be counterintuitive. If there is a strong correlation between two events without any causal connection, sometimes we should allow the inference prohibited by Pearl's reasoning system. For example, suppose that the database has the sentence $\forall x. Take\_cs221(x) \rightarrow_E Brown\_Hair(x)$ together with $C(Take\_cs221(John))$; we can easily imagine a situation in which John is taking cs221 to fulfill his course requirements, hence $Take\_cs221(John)$ is "causally established." It is apparent that we should conclude $Brown\_Hair(John)$. We can justify the reasoning system presented above only when we assume the coexistence of both causal and evidential rules.

With these assumptions in mind, let us look at the following example, which is slightly modified from [Pearl, 1988a].

Example 1
Suppose we have the knowledge about causation that rain causes grass to be wet, that a sprinkler also causes wet grass and that rain causes slippery roads. Its translation into Pearl's representation will be as follows:

$Rain \rightarrow_C WetGrass;$
$Sprinkler \rightarrow_C WetGrass;$
$Rain \rightarrow_C SlipperyRoad;$
$WetGrass \rightarrow_E Rain;$
$WetGrass \rightarrow_E Sprinkler;$
$SlipperyRoad \rightarrow_E Rain.$

If we came to know *WetGrass* because of *Rain*, *WetGrass* would be labeled C according to the assumption mentioned above. When *WetGrass* is labeled C, it cannot be used to infer *Sprinkler* together with the evidential rule $WetGrass \rightarrow_E Sprinkler$, because it is not sanctioned by the system. This shows how two causes of a single effect work *against* each other. On the other hand, if *SlipperyRoad* were labeled E, we would deduce $E(Rain)$ using the evidential rule $SlipperyRoad \rightarrow_E Rain$, and $C(WetGrass)$ would follow from $E(Rain)$ and $Rain \rightarrow_C WetGrass$. In this case, two consequences of a single cause work *for* each other.

3 Difficulty with Pearl's System

With the assumptions mentioned in the previous section in mind, we will now see what can go wrong with Pearl's reasoning system. For instance, when *WetGrass* is established evidentially, it can be used to infer both *Rain* and *Sprinkler*, which is rather counterintuitive. Instead, what we would like to conclude from *WetGrass* is *Rain or Sprinkler*.

One way to overcome this counterintuitive result is to weaken the meaning of the labels (C and E) from that of *acceptance* to that of *support*. That is, we reinterpret the meaning of $E(Sprinkler)$ as *there is evidential support for Sprinkler*, not as *Sprinkler holds for an evidential reason*. Once we reinterpret the E label as support, concluding $E(Rain)$ and $E(Sprinkler)$ from *WetGrass* is no longer counterintuitive. We have reason to believe that it might have rained, and also that the sprinkler might have been on.

Once we weaken the meaning of the E label from acceptance to support, we must also weaken the meaning of the C label. Since one of the three inference rules allows us to infer $C(Q)$ from $E(P)$ and $P \rightarrow_C Q$ and $E(P)$ means only that P is evidentially supported, we are no longer justified in saying that Q holds for a causal reason. Instead, we can say that Q is *causally supported*.

Having a notion of support and making a distinction between causal and evidential support may be useful. However, we still need a mechanism to draw conclusions. If we know it rains, it is clear that we should conclude that the grass is wet, in addition to concluding that there is causal support for the wet grass. In the following sections, we will show how a simple circumscriptive theory can be used to draw conclusions and then see how we can use this circumscriptive theory to define the notion of support.

4 Circumscription for Minimization of Causes

In this section, we propose a simple circumscriptive theory which allows us to draw explanatory conclusions by minimizing causes.

In order to minimize causes, we will reify causal information by using the predicate $causes(P, Q)$. The intended meaning of $causes(P, Q)$ is that P causes Q to hold. We will also use the predicates $holds(P)$ and $holds\_acausally(Q)$, meaning that P is true and that Q is true without any cause being known, respectively. Let the causal theory be divided into two parts, $\mathcal{T} =< R, F >$. R consists of instances of the *causes* predicate together with the following axiom:

$$\forall x.holds(x) \equiv [\exists y.causes(y,x) \wedge holds(y)] \vee holds\_acausally(x) \quad (1)$$

F consists of instances of the *holds* predicate[1].

Given a two-part background theory $< R, F >$, we will circumscribe it by minimizing *causes* and *holds_acausally* with *causes* given higher priority. *causes* is given higher priority because we would like to say that something holds acausally only when we cannot find a cause for it from all we know. That is, we are justified in saying that an event holds acausally only when we do not have any information about its cause.

Example 2
Let our background theory be as follows:

$R : \{causes(Rain, WetGrass),$
$\quad causes(Rain, SlipperyRoad),$
$\quad causes(Sprinkler, WetGrass)\}$ with Axiom (1)
$F : \{holds(WetGrass)\}$

If we circumscribe the background theory in this example, we can conclude $holds(Rain) \vee holds(Sprinkler)$. Consider all the minimal models that satisfy the result of circumscription. In all minimal models, nothing holds acausally. By the axiom (1), we know that $holds(WetGrass)$ is true if and only if any of its causes holds. In this example, the only causes for *WetGrass* are *Rain* and *Sprinkler*. Hence $holds(Rain) \vee holds(Sprinkler)$ is true in all minimal models. However, we will not have $holds(Rain) \wedge holds(Sprinkler)$.

Using the circumscribed background theory, we will define the notion of support in the following section.

5 Supports

As we saw in the previous section, circumscription allows us to draw explanatory conclusions without any unintuitive behavior, and that without the burden of specifying propositional labels initially. However, the notion of causal/evidential support may be useful for certain problems. For instance, we may want to know how one event causally/evidentially affects another even if this event does not logically follow from the other.

As a logical abstraction of probabilistic analysis, we will use a *justification-based* notion of support. First, we will define what a justification is. Then we will define causal and evidential support in terms of justification change and compare our notion of support with the propositional labels in Pearl's $C - E$ system.

5.1 Justification

Informally, a justification for a certain proposition is a *reason to believe* that proposition. Let $\mathcal{T}$ be our background theory. We will define $J_{\mathcal{T}}$ to be a mapping from a well-formed formula to a well-formed formula. If $J_{\mathcal{T}}(\alpha) = \beta$, then α will be true whenever β (a justification for α) is true in the models of the theory $\mathcal{T}$. That is, $\mathcal{T} \cup \{\beta\} \models \alpha$. A formal definition for a justification follows.

Definition 5.1.1 (Justification):
Given a set of first-order sentences $\mathcal{T}$ and a well-formed formula α, $J_{\mathcal{T}}(\alpha) = \beta$ if and only if

$$\beta = \bigvee_i \beta_i$$

for each β_j such that
(1) $\mathcal{T} \cup \{\beta_j\}$ is satisfiable;
(2) $\mathcal{T} \cup \{\beta_j\} \models \alpha$; and
(3) β_j is a conjunction of literals.

Our definition of justification is closely connected with the definition of *minimal support* in the Clause Maintenance System (CMS) by Reiter and deKleer [Reiter and de Kleer, 1987]. They define support as a set of literals which satisfies the conditions (1) and (2) in Definition 5.1.1, and minimal support as a minimal such set. Here we obtain minimality of justification by taking a disjunction of all β_i's rather than requiring each β_i to be minimal. That is, we would like to think of justification as a well-formed formula in disjunctive normal form. Viewing justification as a formula allows us to handle disjunctive explanation easily, thus giving us much more flexibility in defining support.

The third condition deserves some attention. Let us see what happens if we don't have this condition. Given an empty background theory, the justification for P will be P itself. That is, it can only be self-justified because we do not know anything about P. Once we add Q, which may have nothing to do with P, justification for P will change to $\neg Q \vee P$. This is equivalent to $Q \supset P$. This is undesirable because Q can be a random proposition which may have no relevance to P. If we don't have the condition (3), $\tau \supset \alpha$ will be a valid β_i for justification of α for any τ in $\mathcal{T}$.

The addition of condition (3) creates another interesting effect. Suppose our initial background theory was $R \equiv (\neg Q \vee P)$, i.e., we explicitly name $Q \supset P$ as R. Now $J_{\mathcal{T}}(P) = R \wedge Q$. Simply giving a name for $Q \supset P$ causes it to become a part of justification for P. At first glance, it looks rather strange, but in a sense we gave a possibility of using certain literals, in this example R and Q, to express justifications by mentioning them in the background theory. Once we note

[1] How to axiomatize causality is an important problem in and of itself, but it is not what we are interested in. We are interested in how to obtain the proper interaction between causal and evidential support. We believe that the results in the following sections apply independently of the axiomatization used for a causal theory.

this feature, we can use it as a guide for characterizing the terms in which we should express justifications.

Now we will define a partial order on justifications in terms of entailment.

Definition 5.1.2 (Ordering on Justifications): Given two well-formed formulas β_1 and β_2,

(1) $\beta_1 \leq \beta_2$ if and only if $\beta_1 \models \beta_2$, and
(2) $\beta_1 < \beta_2$ if and only if $\beta_1 \leq \beta_2$ and $\beta_2 \not\leq \beta_1$.

Given two different justifications for α, β_1 and β_2, we will say that β_2 is *better* than β_1 if $\beta_1 < \beta_2$, that is, we will say that a semantically weaker justification is better. Essentially a justification for α is a formula we have to add to the background theory so that we can deduce α. If one justification is semantically weaker than others, it means that what we must have in addition to the background theory is weaker. That's why we like a semantically weaker justification better.

We will say that a proposition α is supported when its justification gets better due to the addition of some other proposition β [2]. In the next section, we will define support in terms of an increase in justification, and see how our notion of support is different from Pearl's propositional labels.

5.2 Causal and Evidential Supports

Now we will define causal and evidential supports. In the following, circumscription minimizes *causes* and *holds_acausally* with *causes* given higher priority.

Definition 5.2.1 (Evidential Support): A proposition α is *evidentially* supported with respect to a background theory $\mathcal{T} =< R, F >$ if and only if there is a formula β such that

(1) $\mathcal{T} \models \beta$;
(2) $J_R(\alpha) \not< J_{R\cup\{\beta\}}(\alpha)$;
(3) $J_{CIRC(R\cup F-\{\beta\})}(\alpha) < J_{CIRC(R\cup F)}(\alpha)$.

Definition 5.2.1 says that β evidentially supports α when there is no causal relation between β and α (the second clause), but rather there is evidential relation between them so that the justification for α increases due to circumscription, but not due to the initial background theory.

Definition 5.2.2 (Causal Support): A proposition α is *causally* supported with respect to a background theory $\mathcal{T} =< R, F >$ if and only if there is a formula β such that

(1) β is supported (either evidentially or causally) or $F \models \beta$;
(2) $J_R(\alpha) < J_{R\cup\{\beta\}}(\alpha)$;
(3) $\beta \not\models \alpha$.

β causally supports α when β itself is supported and there is a causal relation between β and α. The third clause is to prevent a circular definition of support. Without it, α may be causally supported if it is either evidentially or causally supported.

Let us work through some examples. Let our background theory have the same R as in Example 2. When F is $\{holds(SlipperyRoad)\}$, $holds(Rain)$ is evidentially supported. In order to see this, we have to compare the justifications for $holds(Rain)$ with and without $holds(SlipperyRoad)$, given a circumscribed background theory. The following relation holds between two different justifications:

$$J_{CIRC(R\cup F-\{holds(SlipperyRoad)\})}(holds(Rain)) < J_{CIRC(R\cup F)}(holds(Rain))$$

Without $holds(SlipperyRoad)$ in the background theory, the justification for $holds(Rain)$ with respect to the circumscribed theory is *unknown*[3]. With $holds(SlipperyRoad)$ in the background theory, the result of circumscription makes the justification for $holds(Rain)$ $True$. Hence we can say that $holds(SlipperyRoad)$ makes the justification for $holds(Rain)$ increase. Also with this background theory, $holds(WetGrass)$ is causally supported because $holds(Rain)$ is evidentially supported and $causes(Rain, WetGrass)$. Here two consequences of a common cause $holds(Rain)$ support each other.

When F is $\{holds(Rain)\}$, $holds(Sprinkler)$ is not evidentially supported because its justification does not increase but rather decreases. Justifications for $holds(Sprinkler)$ with respect to the circumscribed background theory with and without $holds(Rain)$ are as follows:

$$J_{CIRC(R\cup F-\{holds(Rain)\})}(holds(Sprinkler)) = holds(WetGrass) \land \neg holds(Rain)$$
$$J_{CIRC(R\cup F)}(holds(Sprinkler)) = unknown$$

This situation arises because if we apply circumscription to the background theory without $holds(Rain)$, we can infer $holds(WetGrass) \supset holds(Rain) \lor holds(Sprinkler)$. With $holds(Rain)$ in the background theory, however, the justification for $holds(Sprinkler)$ becomes *unknown*. Hence, the justification for $holds(Sprinkler)$ decreases due to the addition of $holds(Rain)$ to the background theory. In this case, two causes of an observation do not support

[2] We were inspired by Gardenfors' work on explanation [Gardenfors, 1988] in that we require an increase in justification, hence a decrease in *surprise*, for a proposition to be supported.

[3] When a proposition can only be self-justified, we will refer to its justification as *unknown*.

each other but rather compete with each other.

Our definition of support is different from Pearl's propositional labels in two ways. The first difference is in its partiality. That is, when a conjunction of multiple events E_1, E_2, and E_3 causes another event E_4, any combination of E_1, E_2, E_3 will causally support E_4; this differs from Pearl's system, in which all three events must happen in order to causally support E_4. As for evidential support, Pearl's system is already partial but there still is a subtle difference. To illustrate this, let R be as in Example 2 and F be $\{holds(WetGrass), holds(SlipperyRoad)\}$. According to definition 5.2.1, $holds(Rain)$ is evidentially supported, but $holds(Sprinkler)$ is not. In Pearl's system $E(Sprinkler)$ will follow if $WetGrass$ and $SlipperyRoad$ are evidentially supported. We believe that this is another aspect of causal reasoning that a logical framework should capture (i.e., when a certain cause must hold in order to explain multiple evidence, it *explains away* others). There is one more difference in the way we define support for events in F. For instance, when our background theory has $R : \{causes(P, Q)\}, F : \{holds(Q)\}$, $holds(P)$ is evidentially supported and $holds(Q)$ is causally supported. However, in Pearl's system, neither $E(P)$ nor $C(Q)$ follows from the theory, $\{P \rightarrow_C Q, Q\}$. This again shows the advantage of our approach.

6 Conclusion

We have presented a very simple circumscriptive theory to draw explanatory conclusions from a causal background theory. We have also showed that we can define the notion of causal/evidential support using circumscription and changes in justification. Not only does our approach avoid certain counterintuitive results, but it also serves as a better logical abstraction of the probabilistic account of causality in the sense that various desirable interactions between causal and evidential supports fall out naturally from their semantic definitions. Our approach provides a notion of both acceptance and support of a proposition. It also allows us to define support without any initial labeling of propositions in the background theory. Finally, it handles the phenomenon of causal asymmetry in a more sophisticated way in that one cause explains away others not only when it is directly known, but also when it can be inferred indirectly from some other propositions. Our approach is not as detailed as a probabilistic analysis, but it is an improvement over Pearl's system.

Acknowledgement

The author would like to thank Matt Ginsberg for his interest, criticism and technical help. Thanks also to Kurt Konolige, Karen Myers, Nils Nilsson, and Martha Pollack for valuable discussions. This work has been supported by NSF under grant IRI-89-12188, by NASA under the research contract NCC2-494, and by General Dynamics.

References

[Gardenfors, 1988] Peter Gardenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Cambridge, Massachusetts, 1988.

[McCarthy, 1986] John McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[Pearl, 1988a] Judea Pearl. Embracing causality in default reasoning. *Artificial Intelligence*, 35:259–271, 1988.

[Pearl, 1988b] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, California, 1988.

[Reiter and de Kleer, 1987] Raymond Reiter and Johan de Kleer. Foundations of assumption-based truth maintenance systems: Preliminary report. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, pages 183–188, 1987.