Causal Reasoning with Domain Constraints

Blai Bonet and Hector Geffner
Departamento de Computación
Universidad Simón Bolívar
Aptdo 89000, Caracas 1080-A
Venezuela
{bonet,hector}@usb.ve

June 20, 1995

Abstract

Default rules are common in everyday discourse yet it has been difficult to incorporate them in AI representation languages. This requires a semantics able to reflect the implicit preferences among defaults and efficient inference algorithms. The interpretation of causal default rules in terms of Qualitative Bayesian Networks (QBN's) [5] is a significant step in this direction. Yet QBN's have limited expressive power and cannot accommodate domain constraints in a natural way. In this work, we develop an interpretation of causal rules, related to QBN's, which accommodates domain constraints. We also present a tractable inference procedure for the resulting system and assess its soundness and completeness properties.
1 Motivation

Default rules are common in everyday discourse yet it has been difficult to incorporate them in AI representation languages. This requires a semantics able to reflect the implicit preferences among defaults and efficient inference algorithms. Problems arise, in particular, when some of the rules carry causal information. Two such rules may be, for example, that a car normally starts when the ignition key is turned but that the car does not start when the battery is dead. Then, in the presence of a reason indicating that the battery is likely to be dead (e.g. the head lights were left on) the car shouldn’t be expected to start. Yet a semantics that simply minimizes default violation yields a different scenario as well where the reason indicating that the battery is dead is suspect (the example is similar to the so-called Yale Shooting Problem discussed in [6]).

Fortunately, the semantical problems that arise from the presence of causal default rules have been nicely tackled in a recent paper by Goldszmidt and Pearl [5]. Their idea is very simple and is based on establishing a correspondence between the causal graph underlying a causal rule system (CRS)—where for every rule there is a directed arc from every variable in the rule antecedent to the (single) variable in the rule consequent—and the causal graph underlying Causal Probabilistic Networks or Bayesian Networks [10]. Provided with the causal graph underlying a CRS, Goldszmidt and Pearl consider the rankings on worlds, in which, like in Bayesian Networks, every variable (literal) is independent of its non-descendants given its parents. Since the rankings can be regarded as order-of-magnitude approximations of probabilities, the resulting semantic structures are very much like an order-of-magnitude approximation of Bayesian Networks, which they call Qualitative Bayesian Networks (QBN’s).

Qualitative Bayesian Networks provide a clear and well-founded interpretation of systems made up of causal rules. Nonetheless they have also significant restrictions that limit their potential use. In particular, QBN’s are not well suited for representing non-causal relations like so-called domain constraints: e.g. that two objects cannot be in the same place
at the same time, or that if I’m in Caracas, I’m in Venezuela, or that death and alive are complementary. *Domain constraints are different than both observations and causal rules.* Domain constraints unlike *observations*, represents assertions that are true not only in the situation at hand but in every conceivable situation. Likewise, domain constraints unlike *causal rules*, establish a relation among variables that is completely symmetric. For example, the rules ‘if the heater is on, the room is hot’ and ‘if the temperature is high, the room is hot’ express a causal relation and a domain constraint respectively. Although the rules are syntactic similar they yield different behavior in the presence of exogenous interventions. For instance, if the window is opened in a freezing night, it is appropriate to contrapose the rule expressing the domain constraint to infer ‘the temperature is not high’, but not the one expressing the causal relation to conclude ‘the heater is off’ (see [11] for an interpretation of causal relations in terms of their behavior in the presence of exogenous interventions). Thus, even while *causal rules, domain constraints and observations may all express constraints and may even correspond to the same syntactic objects (e.g. material implications) they must be given different semantic interpretations*¹.

Bayesian Networks and Qualitative Bayesian Networks accommodate the distinction between causal relations and observations in a natural way: the former are represented by the causal graph and are interpreted as encoding potential independencies; the latter are represented as the value of variables in the graph and enabling or disabling independencies. Domain constraints, on the other hand, are accommodated in a limited way. Basically only those domain constraints that involve a single variable can be expressed.

This led us to ask: is it possible an interpretation of causal rule systems in line with the Bayesian Network reading of causal relations, which can accommodate domain constraints? In this paper we answer this question by constructing such interpretation. The proposed interpretation builds on semantic structures that are related to QBN’s but which differ from QBN’s in significant ways. They are the type of structures used for modelling dynamic

¹Discussion of these distinctions can be founded in [12, 1, 13, 11, 8]
systems. We sketch the intuition and basis of these structures in the next section.

2 Background

Discrete dynamic systems are characterized by a state-transition function that maps states $s_i$ at time $i$, into successor states $s_{i+1}$ at time $i + 1$ [9]. There is sometimes an input function affecting those transitions but we won’t consider them here. In [2] we analyzed the semantics of temporal default theories in terms of slightly different transition function $f$ that map states $s_i$ into sets $f(s_i)$ of possible successor states $s_{i+1}$. With the proper mapping from the theory to the function $f$ we obtained a semantics equivalent to Sandewal’s form of chronological minimization [13] (see also [4]). We also introduced a plausibility function $\pi$ assigning a non-negative integer plausibility rank to each state $s_{i+1}$ so that the plausibilities of the transitions from one state $s_i$ to each of the possible successor states $s_{i+1}$ in $f(s_i)$ are proportional to $\pi(s_{i+1})$. The semantic structures $(f, \pi)$ define dynamic processes analogous to the so-called Markov Chains [7]. When $\pi$ is the uniform plausibility function, i.e. it assigns the same number to all the states, we get the same result as when we only have the transition function $f$.

Dynamic systems involve the notion of time which is not explicitly present in causal theories. In [3], however, we showed that the semantic structures $(f, \pi)$ used for interpreting temporal theories can be extended to deal with causal theories by assigning a ‘time’ point to each variable in the theory, the only restriction to this assignment is that causes precede their effects. In general, there are many possible assignments of that sort, which we represent by ranking functions $\sigma$. The interesting result is that the resulting semantics is insensitive to the function $\sigma$ as long as the transition and plausibility functions $f$ and $\pi$ exhibit certain modularity conditions. Moreover, the resulting modular structures $\langle \sigma, f, \pi \rangle$ are in one to one correspondence with Qualitative Bayesian Networks, and indeed, as in Bayesian Networks parents in the causal graph shield their sons from non-descendants.

In this paper, we extend this results to a language that includes domain constraints,
and show that domain constraints do not impose a burden on the user nor in the processing algorithms.

3 Causal Rule Systems

3.1 Language

We consider causal rule systems (CRS's) comprised of a finite set of causal rules $R$, a finite set of observations $O$, and a finite set of domain constraints $D$. Rules can be either strict or defeasible and will be represented by expressions of the form $\alpha_1 \land \alpha_2 \land \cdots \land \alpha_n \Rightarrow \beta$, and $\alpha_1 \land \alpha_2 \land \cdots \land \alpha_n \rightarrow \beta$ respectively, with $n \geq 1$. The $\alpha$'s and the $\beta$'s are all literals, either negative or positive. Likewise, observations are ground literals, and domain constraints are arbitrary clauses. Rules and domain constraints involving variables stand for the collection of ground instances obtained by replacing the variables from ground terms chosen from a predetermined domain. Thus, a rule like $p(t) \rightarrow q(t + 1)$, where $t$ is a time variable with domain $T = [0, 1000]$, will stand for the collection of rules $p(0) \rightarrow q(1)$, $p(1) \rightarrow q(2)$, \ldots, $p(1000) \rightarrow q(1001)$.

Each defeasible rule in $R$ has a priority number represented by a non-negative integer. The idea is that among a set of incompatible rules, the rules with higher priority are preferred. Unless otherwise stated, all priorities are assumed equal.

Like in Bayesian Networks, we will speak often of variables instead of propositions. So, two literals $a$ and $\neg a$ will be associated with the two possible truth values of an abstract variable $A$.

Finally, we assume that causal systems are acyclic. We define the causal graph of a CRS as a directed graph whose nodes are the variables and where for every causal rule there is a c-link connecting every variable in the antecedent to the variable in the consequent, and for every domain constraint, there are $k$-links connecting all the variables involved in the constraint. A causal cycle in the graph is any directed cycle in the graph involving at least one c-link. A CRS is acyclic iff its causal graph does not contain causal cycles.
3.2 Semantics

As discussed above, we will analyze the semantics of causal rule system in terms of semantic structures of the form \( (\sigma, f, \pi) \), where \( \sigma \) is a ranking function that assigns a non-negative integer \( \sigma(A) \) to every variable \( A \), \( f \) is a state transition function, and \( \pi \) is a plausibility function.

The ranking function \( \sigma \) partitions the set of variables into layers 0, 1, etc., up to a highest layer \( N_\sigma \). The layer \( i \) is comprised of the variables \( A \) such that \( \sigma(A) = i \). No layer can be empty. We will refer to the valuations of the variables in layer \( i \), as local states \( s_i \), and to the valuations of the variables in layers up to and including \( i \), as global states \( s_i \). The transition function \( f \) maps a global state \( s_i \) into global states \( s_{i+1} \) that extend \( s_i \), i.e. \( s_{i+1} = s_i, s_{i+1} \).

We require the ranking \( \sigma \) to satisfy two conditions: first, causes must precede their effects; i.e. if a variable \( A \) occurs in the antecedent of a rule whose head refers to \( B \), then \( \sigma(A) < \sigma(B) \) has to hold. Second, variables involved in the same domain constraint do not have precedence over each other; i.e. for any two variables \( A \) and \( B \) connected by \( k \)-links in the causal graph, \( \sigma(A) = \sigma(B) \) has to hold. Since the theories under consideration are causally acyclic, the existence of such rankings is guaranteed.

The transition and plausibility functions \( f \) and \( \pi \) depend on the structure of the state-space and thus on the ranking \( \sigma \) used. The transition function \( f \) determines the possible successor states \( s_{i+1} \) of a given state \( s_i \). The plausibility function \( \pi \), in the other hand, determines the relative plausibility \( \kappa(s_{i+1} \mid s_i ; \sigma) \) of those possible transitions:

\[
\kappa(s_{i+1} \mid s_i ; \sigma) \overset{\text{def}}{=} \pi(s_{i+1} ; \sigma) - \pi(f^*(s_i ; \sigma) ; \sigma) \quad \text{if } i \geq 0 \text{ and } s_{i+1} \in f(s_i ; \sigma) \quad (1)
\]

Here \( \pi(f^*(s_i ; \sigma) ; \sigma) \) is a normalizing constant that stands for the plausibility of the most plausible local state \( s_{i+1} \) (i.e., with lowest \( \pi \)) such that \( s_i, s_{i+1} \in f(s_i ; \sigma) \). When \( i = 0 \), we define \( \kappa(s_0 ; \sigma) \overset{\text{def}}{=} \pi(s_0 ; \sigma) - \pi(s_0^* ; \sigma) \), where \( s_0^* \) stands for the most plausible state \( s_0^* \). When \( i \geq 0 \), but \( s_i \notin f(s_{i-1} ; \sigma), \kappa(s_i \mid s_{i-1} ; \sigma) = \infty \).
The $\kappa$ plausibility measures can be thought of as an order-of-magnitude approximations of true probabilities as in [5]. Thus the plausibility $\kappa(\tau; \sigma)$ of a trajectory $\tau = s_0, s_1, \ldots, s_{N_\sigma}$ is:

$$\kappa(s_0, s_1, \ldots, s_{N_\sigma}; \sigma) \overset{\text{def}}{=} \kappa(s_0; \sigma) + \sum_{i=0}^{N_\sigma-1} \kappa(s_{i+1}|s_i; \sigma)$$

(2)

Finally, the plausibility of a formula $A$ is defined as $\kappa(A; \sigma) \overset{\text{def}}{=} \min_{\kappa} \kappa(\tau; \sigma)$, and the plausibility of $A$ conditionalized on a set of observations $O$ as $\kappa(A|O; \sigma) \overset{\text{def}}{=} \kappa(A \land O; \sigma) - \kappa(O; \sigma)$.

A different measure for the plausibility of $A$ is obtained when the observations are assimilated by the updating method described in [2, 3]. Both methods will be used in Section 5.

The (legal) trajectories associated with a structure $\langle \sigma, f, \pi \rangle$ will be the sequence of global states $s_0, s_1, \ldots, s_{N_\sigma}$ such that $s_i \in f(s_{i-1}; \sigma)$ for $0 \leq i \leq N_\sigma$. A formula $A$ follows from a set of observations, given $f$ and $\pi$, if the formula is true in all the most plausible trajectories given the observations.

### 3.2.1 Modular Structures

Since alternative rankings $\sigma, \sigma', \ldots$ do not carry relevant information, we want two structures $\langle \sigma, f, \pi \rangle$ and $\langle \sigma', f, \pi \rangle$ sharing the same transition and plausibility functions to behave in the same way. In order to achieve this, we will need certain conditions on $f$ and $\pi$.

Let us refer to the maximal sets of variables connected by $k$-links in the causal graph, as clusters denoted by $\tilde{X}, \tilde{X}', \ldots$. Then, in analogy to [3], we say that $f$ and $\pi$ are decomposable when the past makes the clusters in present independent of each other. This translates, for $\pi$ and $f$, in:

$$\pi(s_i; \sigma) = \sum \pi(\tilde{X}; \sigma) \quad \text{for all } \tilde{X} \in s_i$$

(3)

$$f^L(s_i; \sigma) = \times f(\tilde{X}; s_i; \sigma) \quad \text{for all } \tilde{X} \text{ in layer } i + 1$$

(4)

where $f^L(s_i; \sigma)$ stands for the set of local states $s_{i+1}$ that extends $s_i$ according to $f$, and $f(\tilde{X}; s_i; \sigma)$ stands for the projection of $f(s_i; \sigma)$ over the cluster $\tilde{X}$.

---

2 Variables which do not belong to such clusters are assumed to form singleton clusters.
Likewise, the transition function is local if the past influences the present through the applicables rules only. Formally, if \( \Delta_X(s_i) \) denotes the set of rules whose heads refer to variables in \( \tilde{X} \) and whose body is true in \( s_i \), then \( f \) is local when:

\[
   f_X(s_i; \sigma) = f_X(s_j; \sigma') \quad \text{if} \quad \Delta_X(s_i) = \Delta_X(s_j)
\]

We call the structures \( \langle \sigma, f, \pi \rangle \) in which \( f \) and \( \pi \) satisfies the decomposability and locality conditions modular structures. For modular structures we have results analogous to QBN's:

**Theorem 1** Any two modular structures sharing the same transition and plausibility functions generate the same behavior.

**Theorem 2** In modular structures, the extended parents of a variable in the causal graph shield the variable from all its non-descendants.

Here, the extended parents of a variable \( A \) refer to the union of the parents of all variables in \( A \)'s cluster.

### 3.2.2 Standard Transition Functions

We haven't said yet how to get the transition functions from the rules of the theory. Let \( \Delta(i; \sigma) \) denote the rules that are relevant to the transition from time \( i \) to \( i + 1 \) given the ranking \( \sigma \); i.e. the rules whose head refers to variables in \( \sigma(i + 1) \). We say that \( f(s_i; \sigma) \) is compatible with a subset \( \Delta' \) of \( \Delta(i; \sigma) \) when \( f(s_i; \sigma) \) is the non-empty set of states \( s_{i+1} \) that extend \( s_i \) and satisfy all the rules in \( \Delta' \) as well as the domain constraints. We define the standard transition functions simply as the functions that minimize default violations:

**Definition 1** \( f_s \) is a standard transition function associated with a given set of rules if for every state \( s_i \) and ranking \( \sigma \), \( f_s(s_i; \sigma) \) is compatible with all the strict rules in \( \Delta(i; \sigma) \) and with a lexicographically maximal set of defeasible rules in \( \Delta(i; \sigma) \).

\footnote{By lexicographically maximal we mean a criterion in which all rules with the highest priority are considered first, then if tied, all rules with the second highest priority are considered later and so on. If all rules have the same priorities, lexicographically maximal means just maximal.}
When the system is weakly deterministic\textsuperscript{4}, the standard transition function $f_s$ is unique a satisfies the locality and decomposability conditions. For non-weakly deterministic systems, the same conditions can be obtained with a slightly more complex definition [3]. CRS will be interpreted by structures of the form $\langle \sigma, f_s, \pi \rangle$, for some selected plausibility functions $\pi$.

4 Examples

We are now ready to work some examples. The first addresses some of the classical problems in reasoning about action: the ramification and frame problems. Imagine a box with a red ball inside it. If the box moves then we want to infer the ball moves but that its color doesn’t change, by given a concise representation of the actions effects. For the box and ball problem, the rules and domain constraints are:

\[
\begin{align*}
  r_1 : & \text{Loc}(o,t,l) \rightarrow \text{Loc}(o,t+1,l) \\
  r_2 : & \text{Color}(o,c,t) \rightarrow \text{Color}(o,c,t+1) \\
  r_3 : & \text{Move}(o,l,t) \rightarrow \text{Loc}(o,l,t+1)
\end{align*}
\]

where the variable $c$ ranges over colors, $o$ over objects, $t$ over time, and $l,l'$ over locations. Rules $r_1$ and $r_2$ express persistence relations while $r_3$ express the immediate consequence of moving the box. We assign $r_3$ a higher priority than $r_1$ to give changes priority over persistences. The domain constraints say that the location of the ball and the box coincides and that an object cannot be at two places at the same time.

Using a structure $\langle \sigma, f_s, \pi_{\text{qua}} \rangle$, where for all atoms $a$, $\pi_{\text{qua}}(\neg a) = 0$ and $\pi_{\text{qua}}(a) > 0$, we get the following trajectory given the observations that the ball is red, the box is in NY, and that the box is moved to LA:

\[
\begin{align*}
  \text{Loc(box, 1, NY)} & \quad \text{Loc(box, 2, LA)} & \quad \text{Color(ball, red, 1)} & \quad \text{Move(box, LA, 1)} \\
  \text{Loc(ball, 1, NY)} & \quad \text{Loc(ball, 2, LA)} & \quad \text{Color(ball, red, 2)}
\end{align*}
\]

Thus, as a side effect, moving the box, makes the ball move but does not change its color.

As mentioned above, domain constraints are different than both observations and causal rules. The following example will illustrate these differences. Peter is organizing a party and

\textsuperscript{4}A theory is weakly deterministic if any minimal set of rules whose consequents are jointly inconsistent with the domain constraints, includes a rule with lower priority than the others.
will invite most of his friends. Joe and Pat are among Peter’s friends. Invited people are expected to attend, yet they won’t go if they are sick:

\[ r_1 : \text{Friend}(x) \rightarrow \text{Invited}(x) \quad r_2 : \text{Invited}(x) \rightarrow \text{Go}(x) \quad r_3 : \text{Sick}(x) \Rightarrow \neg \text{Go}(x) \]

Using the same plausibility function \( \pi_{\text{cua}} \), it is easy to show that Joe and Pat will go to the party. Consider now three different reports: Joe and Pat never go to the same party; Joe and Pat would not both go to this party; and if Joe goes, Pat won’t. The first report corresponds to the domain constraint \( d : \neg (\text{Go}(\text{Joe}) \land \text{Go}(\text{Pat})) \) which yields that both are invited but only one attend. The second report corresponds to an observation which yields that either one goes but the other is sick. Finally, the last report, correspond to a strict causal rule which yields that Joe goes but Pat not.

5 Computation

We present now an algorithm for computing the consequences of predictive\(^5\) theories interpreted by structures \( (\sigma, f_s, \pi_{\text{cua}}) \), where \( \pi_{\text{cua}} \) is any plausibility function as above. In addition, the domain constraints are assumed to be definite or negative clauses and that all rules are defeasible. The algorithm uses the set \( O^+ \) that stands for the positive observations, the function closure(\( S \)) that returns the closure of the set \( S \) by the domain constraints, the predicate consistent-closure(\( A, S \)) which tells if closure([\( S \)],\(^6\)) is consistent, and the function new-applicable-rules(\( R, S \)) that returns the rules that become applicable when \( S \) is known and \( R \) is derived.

Algorithm A

\[
\begin{align*}
\text{ACCEPTED} & := \text{closure}(O^+) \\
\text{RULES} & := \text{rules applicable given ACCEPTED} \\
\text{POSSIBLE} & := \emptyset
\end{align*}
\]

\(^5\)We define a theory to be predictive according to whether the observations are assimilated by condition-\(\text{alization (Section 3.2) or by updating [2, 3]. In the first case, a theory is predictive if all the observations refer to root nodes; in the second case, a theory is predictive if predictions are not refuted by observations (see [2, 3] for details).}

\(^6\)The set \([S]_A\) stands for those atoms in the set \( S \) that are in \( A \)'s cluster.
while RULES not empty do
  SELECTED := rules in RULES with highest precedence
  TENTATIVE := ∅
  RULES := RULES - SELECTED
  while SELECTED not empty do
    pick and remove all highest priority rules $A_i \rightarrow p_i$ from SELECTED
    RULES := RULES + new-applicable-rules($\sum_i p_i$, ACCEPTED + POSSIBLE)
    foreach $A_i \rightarrow p_i$ do
      if $A_i \in$ ACCEPTED then TENTATIVE := TENTATIVE + $\{p_i\}$
      else POSSIBLE := POSSIBLE + $\{p_i\}$
      * foreach $p$ in TENTATIVE do
        if ¬consistent-closure($p, p +$ ACCEPTED) then
          remove $p$ from TENTATIVE
        if consistent-closure($p, ACCEPTED + TENTATIVE + POSSIBLE)$ then
          move $p$ from TENTATIVE to ACCEPTED
          remove rules $B \rightarrow p$ from SELECTED and RULES
        else
          move $p$ from TENTATIVE to POSSIBLE
          remove rules $B \rightarrow p$ from SELECTED and RULES
      ACCEPTED := closure(ACCEPTED)
      POSSIBLE := closure(POSSIBLE)
    end while
  end while
end while

**Theorem 3** The complexity of $A$ is linear in the set of rules. If upon completion, ACCEPTED is compatible with the negative observed literals and the theory is predictive, then $A$ is sound. If, in addition, the theory is weakly deterministic, $A$ is complete.

For the ball and box example, the following table shows a trace of $A$ given the observations $Color(ball, red, 1)$, $Loc(box, 1, NY)$ and $Move(box, LA, 1)$ (Here, the closure of $O^+$ is abbreviated as $O^+$):

<table>
<thead>
<tr>
<th>Steps over *</th>
<th>Atoms Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>ACCEPTED = $O^+$, POSSIBLE = ∅, TENTATIVE = ∅</td>
</tr>
<tr>
<td>1</td>
<td>ACCEPTED = $O^+$, POSSIBLE = ∅, TENTATIVE = ${Loc(box, 2, LA)}$</td>
</tr>
<tr>
<td>2</td>
<td>ACCEPTED = $O^+ \cup {Loc(box, 2, LA), Loc(ball, 2, LA)}$, POSSIBLE = ∅, TENTATIVE = ${Color(ball, red, 2)}$</td>
</tr>
<tr>
<td>End</td>
<td>ACCEPTED = $O^+ \cup {Loc(box, 2, LA), Loc(ball, 2, LA), Color(ball, red, 2)}$, POSSIBLE = ∅</td>
</tr>
</tbody>
</table>
This example represents a predictive and weakly deterministic theory, so the algorithm is sound and complete.

References


