Using domain knowledge to select solutions in abductive diagnosis

van Harmelen, F.A.H.; ten Teije, A.

published in
Proceedings of {ECAI}'94
1994

Proceedings of ECAI'94, pp. 652-656, Amsterdam, August 1994

Using domain knowledge to select solutions in abductive diagnosis

Frank van Harmelen [1] and **Annette ten Teije**[1]

Abstract. This paper presents a novel extension to abductive reasoning in causal nets, namely the use of domain knowledge to select among alternative diagnoses. We describe how preferences among multiple causes of a given state can be expressed in terms of causal nets, and how these preferences can be used to select among alternative diagnoses. We investigate this new extension by proving a number of properties, and show how our preference scheme interacts with conventional ways of choosing among competing diagnoses. Our extension increases the expressive power of causal nets, enjoys a number of desirable properties, and compares favourably with existing proposals for expressing preferential knowledge in causal nets.

1 INTRODUCTION

The standard definition of abduction in causal nets (e.g. [2]) often yields multiple possible explanations of a given set of observations, without any further means of distinguishing between these explanations. In this paper, we show how knowledge required for making such selections can be represented in logic-based causal nets: in a specified context, one causal explanation is preferred over another. Such preferences among different causal explanations on the basis of external conditions are often used in practical diagnostic settings. They can be used to select one among all abductive explanations on the basis of additional domain knowledge about plausibility, danger, urgency, cost etc.

Our particular encoding of such selection knowledge has some desirable properties: the addition of selection knowledge leaves the standard notion of abductive solution unperturbed; certain types of selected solutions are guaranteed to exist whenever a normal solution exists; and standard causal nets are a special case of our extended causal nets.

Section 2 summarises some of the basic definitions from the literature. Section 3 presents our extensions of causal nets with selection conditions and gives some examples and properties. Section 4 discusses the interaction between our selection conditions and conventional, more syntactic ways of selecting among abductive explanations. Section 5 concludes.

2 BASIC DEFINITIONS

In causal nets, states in the world are represented as predicates, and the fact that state S_i necessarily causes state S_j is represented as $S_i \rightarrow S_j$. A weaker notion of causality can be modelled as proposed in [3]: S_i possibly causes S_j (but not necessarily) can be written as $S_i \wedge \alpha_{ij} \rightarrow S_j$, where α_{ij} is interpreted as the unknown condition required for S_i to cause S_j. A further refinement is the notion of context: information about the world that need not itself be causally explained, but that influences causal transitions. If S_i causes S_j in context C_{ij}, we write $C_{ij} \rightarrow (S_i \rightarrow S_j)$. An example of contextual information is that in women, pregnancy causes high blood pressure. Womanhood is not one of the causes of high blood pressure, but is an enabling condition for the causal link between pregnancy and high blood pressure. The crucial aspect of such contextual knowledge is that it can be observed (like symptoms), but that unlike symptoms, it need not be explained by causal chains. Notice that the representation of contextually dependent (or: conditional) causal links is the same as for possible causal links (since $C_{ij} \rightarrow (S_i \rightarrow S_j) \Leftrightarrow (C_{ij} \wedge S_i \rightarrow S_j)$), but their treatment in the definitions of diagnostic problem and solution are different: the truth values of conditions are given, while the value of unknown symbols must be assumed. Obviously, conditional and possible causality can be combined, with $C_{ij} \rightarrow (S_i \wedge \alpha_{ij} \rightarrow S_j)$ stating that under conditions C_{ij}, S_i possibly causes S_j. All this is summarised in the following, more or less standard definition of causal net:

Definition 2.1 (Causal Net) *A* **causal net** $\mathcal{N}$ *is a set of propositional sentences*[2] *of the following form:*

$$\begin{array}{rcll} S_i & \rightarrow & S_j & \textit{(necessary causality)} \\ S_i \wedge \alpha_{ij} & \rightarrow & S_j & \textit{(possible causality)} \\ C_{ij} \rightarrow (S_i & \rightarrow & S_j) & \textit{(conditional causality)} \\ C_{ij} \rightarrow (S_i \wedge \alpha_{ij} & \rightarrow & S_j) & \textit{(conditional possible causality)} \end{array}$$

Furthermore, we define the following terminology:
- INIT($\mathcal{N}$), *the initial nodes, are those* S_i *only occurring on the lefthand side of implications;*
- OBS($\mathcal{N}$), *the terminal nodes, are those* S_i *only occurring on the righthand side of implications;*
- UNKNOWN($\mathcal{N}$) *the unknown symbols, are the* α_{ij}*;*
- COND($\mathcal{N}$), *the conditional symbols, are the* C_{ij}.

These causal nets can be used in diagnostic problem solving. In a diagnostic problem, we are given a causal net, some observations and a context, and we want to find an abductive explanation that in the given context (i) implies the observations (ii) does not imply any of the absent observations, and (iii) is consistent with the causal net.

Definition 2.2 (Diagnostic Problem) *A* **diagnostic problem** *is a tuple* $\langle \mathcal{N}, \mathcal{C}, \mathcal{O}^+ \rangle$*, with* $\mathcal{N}$ *a causal net,* $\mathcal{C} \subseteq$ COND($\mathcal{N}$) *(the true contextual conditions) and* $\mathcal{O}^+ \subseteq$ OBS($\mathcal{N}$) *(the observed symptoms).*

---

[1] SWI, University of Amsterdam, Roetersstraat 15, 1018WB Amsterdam, email: {frankh,annette}@swi.psy.uva.nl

[2] We limit ourselves to representing states by propositional letters, but [2] shows that the extension to the predicate case is unproblematic.

ECAI 94. *11th European Conference on Artificial Intelligence* Edited by A. Cohn
Published in 1994 by John Wiley & Sons, Ltd.

Definition 2.3 (Solution to a Diagnostic Problem) *A* **solution** *to a diagnostic problem* $\langle \mathcal{N}, \mathcal{C}, \mathcal{O}^+ \rangle$ *is a tuple* $\langle \mathcal{I}, \mathcal{A} \rangle$ *with* $\mathcal{I} \subseteq \text{INIT}(\mathcal{N})$ *and* $\mathcal{A} \subseteq \text{UNKNOWN}(\mathcal{N})$ *such that*

$$\mathcal{N} \cup \mathcal{C} \cup \mathcal{I} \cup \mathcal{A} \vdash \mathcal{O}^+ \quad (1)$$

$$\mathcal{N} \cup \mathcal{C} \cup \mathcal{I} \cup \mathcal{A} \cup \mathcal{O}^- \not\vdash \bot^3 \quad (2)$$

Notice that we align ourselves with [3] in demanding that a solution to a diagnostic problem should imply the observed symptoms and be consistent with absent symptoms, although other choices in this matter are possible, as described in [3]

In general, we are interested in the smallest solutions, which explain the observations through a minimal set of abnormal states:

Definition 2.4 (Diagnosis as Minimal Solution) *A solution* $\langle \mathcal{I}, \mathcal{A} \rangle$ *is a* **minimal solution** *(written* $\subset$*-min*$(\langle \mathcal{I}, \mathcal{A} \rangle)$*)* iff *no proper subset of* $\mathcal{I} \cup \mathcal{A}$ *is a solution. Such a minimal solution is called a* diagnosis.

3 SELECTION CONDITIONS

3.1 Why preferences are useful

The standard definition of minimal abductive diagnosis in causal nets as given above often yields multiple explanations for a given set of observations, without any further ways to distinguish between these alternative explanations. In practical diagnostic settings, experts exploit knowledge about the preference for different explanations under different circumstances. Such additional domain knowledge can be used to make a selection among multiple abductive explanations. We might want to choose the most probable explanation, or the most dangerous one (if remedial action must be taken), or the cheapest one to confirm (if further checks must be carried out), etc.

A typical example of such knowledge is: "A non-working light can be explained by either a broken bulb or by a flat battery, but if the temperature is sub-zero, a flat battery is the most likely explanation". Thus, contextual knowledge (such as the temperature being sub-zero) expresses a *preference* among the various possible causes of a state. Practical knowledge-engineering tools like MOLE [7] do indeed exploit such knowledge in the causal nets that they elicit from domain experts. Such knowledge cannot be encoded in the nets from the previous section, since the sub-zero temperature does not by itself *cause* the observation, nor is it a necessary *condition*. It is however very desirable that such knowledge should be represented in causal nets for diagnosis.

- The preferences among different possible causes for a single state reduces the *number* of diagnoses.
- The knowledge can be used to prefer diagnoses with a particular property, since the selection criteria are based on case specific contextual knowledge. There preferences might be based on directly observable circumstances (such as the sub-zero temperature mentioned above), but might also be used to enforce desirable properties of a diagnosis such as urgency or cost.
- The *search* for diagnoses can be performed faster, since during the search, non-preferred causes can be removed from the search-space, and the computation can focus on the preferred causes.

We therefore define a new type of causal net, which enriches the nets from the previous section with expressions which indicate that the causal link between S_i and S_j is preferred under condition P_{ij}.

[3] In this paper, we define $\mathcal{O}^-$ as $\{\neg O \mid O \in OBS(\mathcal{N}) \backslash \mathcal{O}^+\}$, representing the absent symptoms. A straightforward generalisation is to take $\mathcal{O}^-$ as some subset of this. That would allow a distinction between absent symptoms ($\mathcal{O}^-$) and unknown symptoms (namely those symptoms neither in $\mathcal{O}^+$ nor in $\mathcal{O}^-$).

3.2 Soft selections

Notation: We write $S_i \leadsto S_j$ for a causal link of any type from S_i to S_j.

Definition 3.1 (Causal Nets with Selection Conditions)
A **causal net with selection conditions** *is a causal net* $\mathcal{N}$ *as in Def. 2.1, with the following changes:*
- *For every causal link* $S_i \leadsto S_j$ *in* $\mathcal{N}$*, we introduce a* **preference predicate** P_{ij}*. (If absent, such a predicate is assumed to be false).*
- *For every causal link* $S_i \leadsto S_j$ *in* $\mathcal{N}$ *we define the* **selection condition** Π_{ij} *as* $\Pi_{ij} \equiv (P_{ij} \vee \bigwedge_k \neg P_{kj})$.

The P_{ij} are intended to be the domain specific predicates which express when a causal link is preferred over others (such as the sub-zero temperature in the earlier example). The Π_{ij} will capture when a causal link is selected for an explanation: either its preference predicate P_{ij} is true, or it is false, but all of the competing neighbouring links (all $S_k \leadsto S_j$) also have a false preference predicate.

The definitions for diagnostic problem and solution remain almost unchanged for this new type of causal net, so we do not repeat them here. Definition 2.2 of a diagnostic problem must be trivially adapted to become $\langle \mathcal{N}, \mathcal{C}, \mathcal{P}, \mathcal{O}^+ \rangle$ to take into account the given truth-values of the additional contextual symbols P_{ij}. Formula 2 in Definition 2.3 must be trivially changed to include $\mathcal{P}$ in the consistency demand, in order to enforce consistency of $\mathcal{P}$.

In Def. 3.1 a Π_{ij} indicates that the corresponding link $S_i \leadsto S_j$ is selected. Obviously, we want to select those solutions to a diagnostic problem where the selection conditions Π_{ij} have played a maximally guiding role. We encode this in the following use of the Π_{ij}:

Notation: We write Σ_j (defined as $\Sigma_j = \{S_i \leadsto S_j \mid \neg\Pi_{ij}\}$) for the set of unselected links causing S_j. We write Σ (defined as $\Sigma = \bigcup_j \Sigma_j$) for the set of all unselected links in $\mathcal{N}$.

Definition 3.2 (Local Preference) *A solution* $\langle \mathcal{I}, \mathcal{A} \rangle$ *is* **locally preferred** *over* $\langle \mathcal{I}', \mathcal{A}' \rangle$ *(written* $\langle \mathcal{I}, \mathcal{A} \rangle <_{LP} \langle \mathcal{I}', \mathcal{A}' \rangle$*)* iff

$$\begin{array}{llr} \exists j : & \mathcal{N} \backslash \Sigma_j \cup \mathcal{C} \cup \mathcal{I} \cup \mathcal{A} \vdash \mathcal{O}^+ & (3.2.1) \\ & \mathcal{N} \backslash \Sigma_j \cup \mathcal{C} \cup \mathcal{I}' \cup \mathcal{A}' \not\vdash \mathcal{O}^+ & (3.2.2) \\ & \mathcal{N} \cup \mathcal{C} \cup \mathcal{I} \cup \mathcal{A} \vdash S_j & (3.2.3) \\ & \mathcal{N} \cup \mathcal{C} \cup \mathcal{I}' \cup \mathcal{A}' \vdash S_j & (3.2.4) \end{array}$$

The rationale of this is as follows: $\mathcal{N} \backslash \Sigma_j$ represents the causal net with the causes of S_j restricted to only the selected links (since Σ_j, the unselected causes of S_j, are removed). The definition now states that $\langle \mathcal{I}, \mathcal{A} \rangle$ is preferred over $\langle \mathcal{I}', \mathcal{A}' \rangle$ *iff* $\langle \mathcal{I}', \mathcal{A}' \rangle$ uses an unselected link through S_j to imply $\mathcal{O}^+$ (since the removal of the unselected links through S_j prevent $\langle \mathcal{I}', \mathcal{A}' \rangle$ from implying $\mathcal{O}^+$; (3.2.2)), whereas $\langle \mathcal{I}, \mathcal{A} \rangle$ does not use an unselected link through S_j (since it still implies $\mathcal{O}^+$, even after removal of all the unpreferred links through S_j; (3.2.1)). Thus, at location S_j, $\langle \mathcal{I}', \mathcal{A}' \rangle$ makes essential use of an unselected link, whereas $\langle \mathcal{I}, \mathcal{A} \rangle$ does not. The third and fourth condition above ensure that this distinction between $\langle \mathcal{I}, \mathcal{A} \rangle$ and $\langle \mathcal{I}', \mathcal{A}' \rangle$ occurs at a node S_j which does indeed play a role in both solutions (i.e. a node which occurs somewhere on a path in $\mathcal{N}$ between diagnosis and observations).[4]

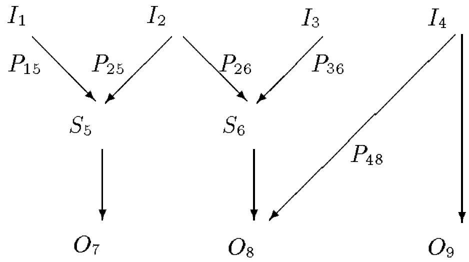
We only have reason to prefer an entire solution Sol_1 over another solution Sol_2 if Sol_1 is somewhere locally preferred over Sol_2, and nowhere else is Sol_2 locally preferred over Sol_1:

[4] Strictly speaking, (3.2.4) is redundant since it is already implied by (3.2.2) plus the fact that $\langle \mathcal{I}', \mathcal{A}' \rangle$ is a solution.

Definition 3.3 (Global Preference) *A solution Sol_1 is* **globally preferred** *over Sol_2 (written $Sol_1 <_{GP} Sol_2$)* iff
$(Sol_1 <_{LP} Sol_2) \wedge \neg(Sol_2 <_{LP} Sol_1)$.

The definition of a selected solution is then as expected:

Definition 3.4 (Selection)
A solution Sol_1 is **selected** *(written $selected(Sol_1)$)* iff *there is no other solution Sol_2 with $Sol_2 <_{GP} Sol_1$.*



	P_{15}	P_{25}	P_{26}	P_{36}	selected solutions				
1	0	0	0	0	$\{I_1, I_2\}$	$\{I_1, I_3\}$	$\{I_2\}$	$\{I_2, I_3\}$	$\{I_1, I_2, I_3\}$
2	0	1	0	0	$\{I_1, I_2\}$		$\{I_2\}$	$\{I_2, I_3\}$	$\{I_1, I_2, I_3\}$
3	1	0	1	0	$\{I_1, I_2\}$				$\{I_1, I_2, I_3\}$
4	1	1	0	0	$\{I_1, I_2\}$	$\{I_1, I_3\}$	$\{I_2\}$	$\{I_2, I_3\}$	$\{I_1, I_2, I_3\}$

Figure 1.

Example 3.5 Figure 1 shows an example net with preference predicates. All links are necessary causal links. The table shows the selected solutions for $\mathcal{O}^+ = \{O_7, O_8\}$ and various truth-values of preference predicates.
• Case 1 in this table has all preferences false, resulting in exactly the normal abductive solutions (according to Definition 2.3), of which $\{I_1, I_3\}$ and $\{I_2\}$ are the $\subset$-min solutions (Definition 2.4).
• Case 2 illustrates the effect of a true preference predicate. Since P_{25} is true, Π_{25} is true and Π_{15} is false. This forces the use of I_2 to explain O_7, thereby removing $\{I_1, I_3\}$ as a solution.
• Case 3 shows a configuration of preference predicates which leads to a greatly reduced set of selected solutions. For instance, solution $\{I_2, I_3\}$ is no longer selected, since it uses the unselected link $I_2 \rightarrow S_5$ while the selected solution $\{I_1, I_2\}$ uses the preferred alternative $I_1 \rightarrow S_5$.
• Case 3 is also an illustration of solutions which are locally preferred but not globally, and are therefore not selected: we have $\{I_1, I_3\} <_{LP} \{I_2\}$, but since we also have $\{I_2\} <_{LP} \{I_1, I_3\}$ neither solution is globally preferred and thus neither is selected.
• Case 4 shows a configuration of P_{ij} which does not lead to a reduction. At every node S_j, either all P_{ij} are true or all P_{ij} are false. This makes every Π_{ij} true, and therefore results in the entire set of solutions being selected, as in case 1.

3.3 Hard selections

In Definitions 3.2–3.4, we have used preference predicates to express a preference among abductive explanations. Under these definitions, unselected links can still be used in an explanation, namely when selected links do not suffice for an explanation. An alternative, and more strict use of the selection criteria would be to demand that only those links can be used in an explanation whose selection condition Π_{ij} is true. We will call this use of the selection criteria *hard selection*, (and refer to the selection from Definition 3.4 as *soft* selection).

Definition 3.6 (Hard Selected)
A solution $\langle \mathcal{I}, \mathcal{A} \rangle$ is **hard selected** *(written hard-selected($\langle \mathcal{I}, \mathcal{A} \rangle$))* iff $\mathcal{N} \backslash \Sigma \cup \mathcal{C} \cup \mathcal{P} \cup \mathcal{I} \cup \mathcal{A} \vdash \mathcal{O}^+$

The rationale of this definition is that it demands of a solution that it is still a solution if we restrict the net to only those links with a true selection condition.

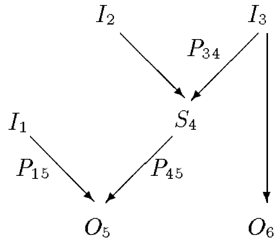


Figure 2.

Example 3.7 *In the net of Figure 2, if $\mathcal{O}^+ = \{O_5\}$ and with P_{15}, P_{45} and P_{34} as the only true preference predicates, then both $\{I_1\}$ and $\{I_2\}$ are soft-selected solutions, whereas only $\{I_1\}$ is a hard-selected solution (since $\{I_2\}$ uses the unpreferred link $I_2 \rightarrow S_4$).*

3.4 Properties

Whereas the truth or falsehood of a condition from $\mathcal{C}$ determines the presence or absence of a causal link in the net, the truth or falsehood of a preference predicate from $\mathcal{P}$ does not affect the presence of the link, but only how this link will be used. This is captured in the following property, which follows from the fact that Definition 2.3 remains essentially unchanged[5]:

Property 3.8 (Selection conditions do not affect solutions)
$\langle \mathcal{I}, \mathcal{A} \rangle$ is a solution for $\langle \mathcal{N}, \mathcal{C}, \mathcal{O}^+ \rangle$ iff *$\langle \mathcal{I}, \mathcal{A} \rangle$ is a solution for $\langle \mathcal{N}, \mathcal{C}, \mathcal{P}, \mathcal{O}^+ \rangle$ for any consistent value of $\mathcal{P}$.*

A desirable property is that a soft-selected diagnosis always exists if a diagnosis exists at all:

Property 3.9 (Existence of soft-selected solutions) *If $\langle \mathcal{N}, \mathcal{C}, \mathcal{O}^+ \rangle$ has a solution then $\langle \mathcal{N}, \mathcal{C}, \mathcal{P}, \mathcal{O}^+ \rangle$ has a soft-selected solution for any $\mathcal{P}$.*

This property fails for hard-selected solutions. This can be seen in the net from figure 1. If we take again $O^+ = \{O_7, O_8\}$, but now we take P_{48} as the only true preference predicate, then no solution exists, since of the three explanations of O_8 (namely I_2, I_3 and I_4), the first two fail the hard selected criterion of definition 3.6, and the last one is disabled because of its inconsistency with the unobserved O_9.
The following property shows that the standard nets from Definition 2.1 are a special case of our extended nets: a net without preference conditions behaves as a net with such conditions if all the P_{ij} are either true or false:

Property 3.10 (Reduction to net without selection conditions)
When either all $P_{ij} \equiv \bot$ or all $P_{ij} \equiv \top$ then every solution is hard-selected (and by property 3.11 also soft-selected).

Hard selected solutions are a special case of soft selected solutions:

Property 3.11 (Inclusion of hard selections in soft selections)
If hard-selected(Sol) then soft-selected(Sol)

[5] Because of lack of space, we omit proofs

Finally, both hard and soft selection turn out to be monotonic in the size of solutions:

Property 3.12 (Monotonicity of selected solutions) *If Sol_1, Sol_2 are both solutions to a diagnostic problem, with $Sol_1 \subseteq Sol_2$ then:*
- *If soft-selected(Sol_1) then soft-selected(Sol_2).*
- *If hard-selected(Sol_1) then hard-selected(Sol_2).*

This means that hard and soft selection are by themselves not sufficient to produce minimal solutions, since every solution containing a selected solution is again selected. We will need to combine our selection criteria with other minimality criteria to achieve satisfactory minimal solutions. This is the subject of the next section.

4 COMBINATION WITH EXISTING MINIMALITY CRITERIA

Property 3.12 shows that it will be necessary to combine our selection criteria with minimality criteria such as $\subset$-min (from Definition 2.4), since selected solutions by themselves are not sufficiently restricted: any solution containing a selected solution is again selected. In this section we investigate how the syntactically oriented minimality criteria such as $\subset$-min interact with our selection criteria.
One of the simplest ways to combine minimality criteria is lexicographically: apply one criterion to solutions minimal under the other. We could for instance apply selection to all $\subset$-min solutions. For this we write soft-$\subset$-min, defined by:

Definition 4.1 (subset-minimality then soft-selection)
For any solution S, soft-$\subset$-min(S) iff
$\subset$-min(S) $\wedge \neg\exists S'$: $\subset$-min(S') $\wedge\ S' <_{GP} S$.

Thus, a solution is soft-$\subset$-min *iff* it is soft-selected among the $\subset$-min solutions. This combination has the pleasant property that the selection criteria further reduce the number of $\subset$-min solutions, and in fact always produce a subset of these:

Property 4.2 (Reduction of diagnoses by selection)
If soft-$\subset$-min(Sol) then $\subset$-min(Sol).

Although trivial, we state this property here, since it shows that the selection conditions (if used in this way) do indeed reduce the number of solutions. This property does not hold when soft selection and $\subset$-min are combined in the opposite order. If we define $\subset$-soft-min as

Definition 4.3 (Soft-selection then subset-minimality)
For any solution S, $\subset$-soft-min(S) iff
soft-selected(S) $\wedge \neg\exists S'$: soft-selected(S') $\wedge\ S' \subset S$.

(i.e. $\subset$-soft-min(S) *iff* S is $\subset$-min among the soft-selected solutions), then the following example illustrates that property 4.2 does not always hold for this combination of minimalities:

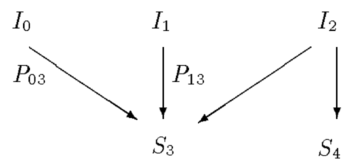


Figure 3.

Example 4.4 *In Figure 3, when $\mathcal{O}^+ = \{S_3, S_4\}$, and P_{03} and P_{13} are true, then using $\subset$-soft-min($\cdot$) gives $\{I_0, I_2\}$ and $\{I_1, I_2\}$ as diagnoses, whereas ignoring the P_{ij}, we would only have $\{I_2\}$ as a $\subset$-minimal diagnosis.*

A third combination the orderings is to apply them simultaneously:

Definition 4.5 (Soft-selection and subset-minimality) *For any solution S, soft-&-$\subset$-min(S)* iff *soft-selected(S) $\wedge$ $\subset$-min(S)*

This is a rather strong minimality criterion, and it does not even guarantee the existence of minimal solutions when solutions exist, a property which does hold for the other two combinations:

Property 4.6 (Existence of minimal solutions)
- *A soft-$\subset$-min and a $\subset$-soft-min solution exist* iff *a solution exists.*
- *A soft-&-$\subset$-min solution is not guaranteed to exist when a solution exists*

Example 4.4 shows the possible non-existence of soft-&-$\subset$-min solutions: $\{I_0, I_2\}$, $\{I_1, I_2\}$ and $\{I_0, I_1, I_2\}$ are soft-selected solutions, while $\{I_2\}$ is the only $\subset$-min solution, so no solution exists which satisfies both criteria.
Since property 3.12 (monotonicity) also holds for hard-selected solutions, we must also combine hard-selected with $\subset$-min to obtain useful minimal hard selected solutions. We can define $\subset$-hard-min and hard-&-$\subset$-min analogously to definitions 4.3 and 4.5. The following gives a full picture of the relationship between these criteria.

Property 4.7 (Increasing chains of selection criteria)

hard-&-$\subset$-min $\rightarrow$ *soft-&-$\subset$-min* $\rightarrow$ <u>*soft-$\subset$-min*</u> $\rightarrow$ <u>*$\subset$-min*</u>
$\searrow$ $\searrow$
$\subset$-hard-min <u>*$\subset$-soft-min*</u> $\rightarrow$ <u>*soft-selected*</u>
$\searrow$ $\nearrow$
hard-selected

This shows that we can combine the various selection mechanisms into ever more selective configurations.The underlined combinations are guaranteed to yield a minimal solution if any solution to the diagnostic problem exists at all. The other combinations may return no solution, even though definition 2.3 would allow one.
Which of these minimality criteria should be chosen depends on application specific criteria. One might want to use the domain specific preference criteria before general measures like $\subset$-min. This would lead us to choose $\subset$-soft-min over soft-$\subset$-min. Secondly, stronger minimality criteria would seem preferable, leading us to criteria toward the left of the picture. This is counteracted by the desire not to loose real solutions through minimality criteria. This would restrict us to the underlined combinations. Another possibility is first to use stronger combinations, and to relax these for those problems where no solution is returned.
It seems that none of these or other combinations of minimality criteria and preferences is ideal for all situations. This is exactly as predicted in [5], where Doyle proves that no combination of preferences exists which is optimal (or even acceptable) under all circumstances.

5 DISCUSSION

Computational issues: Definition 3.6 shows that finding hard-selected solutions corresponds to finding normal solutions in a reduced network (replacing $\mathcal{N}$ by $\mathcal{N}\backslash\Sigma$). Thus, finding hard-selected solutions will in general be easier and certainly not harder than finding normal solutions. Similarly, we can reduce computing soft-selected solutions in $\mathcal{N}$ to computing hard-selected solutions in $\mathcal{N}' = \mathcal{N}\backslash\{S_i | \exists o \in \mathcal{O}^- : \mathcal{N} \cup \mathcal{C} \cup S_i \vdash o\}$. However, because of the complexity of reducing $\mathcal{N}$ to $\mathcal{N}'$ this does not immediately give an efficient algorithm for computing soft-selected solutions.

Comparison with other work: The MOLE system [7] is a practical diagnostic system that uses preferences. However, no declarative account of the preferences is given, and this deficiency was in fact one of the motivations that lead us to this work. It turns out that MOLE's preferences correspond exactly to the hard preferences from Definition 3.6.

The plausabilities in [1] resemble our preferential knowledge, but are only defined for initial nodes, while our preferences can be defined on arbitrary links, anywhere in the net. The same remark holds for the mode preferences from [6].

Although the probabilities from [11] may at first sight resemble preferential knowledge, they are in fact quite different. The probabilities encode the strength of the causal link (which may then be used for preference purposes), whereas our preferences are independent of the strength of the causal link, and can be used to encode preferences on the basis of not only plausibility, but also danger, cost, urgency etc. The same remark holds for [12].

Although not in a diagnostic context, [13] has proposed a scheme for weighted abduction where literals in a theory (in our case: nodes in a causal net) are equipped with numerical weights, and minimisation of weights is used to select among competing explanations. However, assignment of the numerical weights is problematic, and no semantics exists for the interpretation of these weights.

Much work that is close in motivation to our own has been done in the context of consistency-based diagnosis. [8] discusses the use of prioritised defaults to express preferences among diagnoses. The resulting preference mechanism is not very flexible, and comes at the price of a rather non-standard formalism. [6] describes a way of expressing preferences among fault modes which is at the same time more flexible then Junker's and uses a simpler formalism (normal default logic). Our work differs from these approaches in a number of ways: it is set in an abductive rather than a consistency-based context; we use only a standard first order formalism; and both [8] and [6] express only preferences among fault-modes (roughly corresponding to our set INIT($\mathcal{N}$)), whereas we exploit the structure of the entire theory to express preferences.

Our preferences are a way to select among competing solutions. Other work (e.g. [10]) tries to obtain new discriminating observations to select among competing solutions. Our preferences are a selection method that should be applied when it is no longer possible or desirable to obtain further discriminating observations.

Future work: An obvious extension that is required to make the results from this paper more useful is to allow $\mathcal{N}$ to be a Horn Clause theory. The results from [2] lead us to believe that this will not present any significant difficulty.

Extending our results beyond Horn Clause theories might be more problematic. Because we attach the preference predicates to causal links, we make very specific assumptions about the syntactic form of $\mathcal{N}$. Although these assumptions are widely made in the literature on abductive diagnosis [2, 12], other formalisations of abduction deal with $\mathcal{N}$ as an arbitrary theory (e.g. [9], or the abductive case of [4]). It is unclear how our approach could be used in theories that do not impose a particular syntactic form for individual causal links.

A second extension involves an enlarged vocabulary for expressing preferential knowledge. At first sight, it would seem useful to allow also negative preferences ("do *not* use this link under a certain condition"), and preferential knowledge on states as well as on links (this vocabulary is used in MOLE [7]). Preliminary investigations suggest that these extensions can all be expressed in terms of the formalism of this paper.

Achievements: We have extended logic-based causal nets by annotating causal links with selection criteria. These selection criteria provide a way to select abductive explanations on the basis of additional domain knowledge (the preference predicates P_{ij}). This differs from standard techniques in the literature, which are typically very syntactic in nature. They are based on the topology of the net, whereas our selection criteria are based on domain specific knowledge.

The preference conditions can be used in two ways to select among competing abductive solutions: hard and soft selection. We have shown how these nets, and the solutions they produce, are related to the standard causal nets: normal nets are a special case of nets with preference conditions; soft selected solutions are guaranteed to exist when normal solutions exist; hard selected solutions are a special case of soft solutions and are not always guaranteed to exist. Both types of selection suffer from a monotonicity property which requires the combination of the selection scheme with other minimality criteria. We showed how our selection schemes can be combined with such minimality criteria in a large variety of ever more selective configurations. Our representation of preferential knowledge is completely orthogonal to other extensions of causal nets such as conditional causality and possible causality, coherent hypothesis sets, abstraction layers in causal nets etc, enabling an unproblematic combination between our preference conditions and these numerous other extensions.

REFERENCES

[1] T. Bylander, D. Allemang, M. C. Tanner, and J. R. Josephson, 'The computational complexity of abduction', *Artificial Intelligence*, **49**, 25–60, (1991).

[2] L. Console and P. Torasso, 'Hypothetical reasoning in causal models', *Int. J. of Intelligent Systems*, **5**(1), 83–124, (1990).

[3] L. Console and P. Torasso, 'Integrating models of the correct behaviour into abductive diagnosis', in *ECAI'90*, pp. 160–166, (1990).

[4] J. de Kleer, A.K. Mackworth, and R. Reiter, 'Characterizing diagnoses and systems', *Artificial Intelligence*, **56**(2–3), 197–222, (1992).

[5] J. Doyle, 'Impediments to universal preference-based default theories', *Artificial Intelligence*, **49**, 97–127, (1991).

[6] O. Dressler and P. Struss, 'Back to defaults: characterising and computing diagnoses as coherent assumption sets', in *ECAI'92*, pp. 719–723, (1992).

[7] L. Eshelman, 'MOLE: A knowledge-acquisition tool for cover-and-differentiate systems', in *Automating Knowledge Acquisition for Expert Systems*, ed., S. Marcus, 37–80, Kluwer, Boston, (1988).

[8] U. Junker, 'Prioritised defaults: Implementation by TMS and application to diagnosis', in *IJCAI'91*, pp. 310–315, (1991).

[9] K. Konolige, 'Abduction versus closure in causal theories', *Artificial Intelligence*, **53**, 255–272, (1992).

[10] S. McIlraith and R. Reiter, 'On tests for hypothetical reasoning', in *Readings in Model-Based Diagnosis*, eds., W. Hamscher, L. Console, and J. de Kleer, 89–96, Morgan Kaufman, (1992).

[11] J. Pearl, 'Distributed revision of composite beliefs', *Artificial Intelligence*, **33**, 173–215, (1987).

[12] D. Poole, 'Representing diagnostic knowledge for probabilistic Horn abduction', in *IJCAI'91*, pp. 1129–1135, (1991).

[13] M. Stickel, 'Rationale and methods for abductive reasoning in natural-language interpretation', in *Proceedings Natural Language and Logic*, ed., R. Studer, volume 459 of *Lecture Notes in AI*, pp. 233–252, (1989).