Exploiting domain knowledge for approximate diagnosis

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

published in
Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence ({IJCAI}'97)
1997

document version
Publisher's PDF, also known as Version of record

Proceedings of IJCAI'97, pp. 434-439

Exploiting domain knowledge for approximate diagnosis

Annette ten Teije
SWI
University of Amsterdam
annette@swi.psy.uva.nl

Frank van Harmelen
Dept. of Math. and CS
Vrije Universiteit Amsterdam
frankh@cs.vu.nl

Abstract

The AI literature contains many definitions of diagnostic reasoning most of which are defined in terms of the logical entailment relation. We use existing work on approximate entailment to define notions of approximation in diagnosis. We show how such a notion of approximate diagnosis can be exploited in various diagnostic strategies. We illustrate these strategies by performing diagnosis in a small car domain example.

1 Motivation

The AI literature contains many definitions of diagnostic reasoning. However, there are many reasons why we should not search for *the* appropriate definition of diagnosis, but instead search for alternative definitions, and investigate how they relate to each other. There exists a whole space of reasonable notions of diagnosis. These notions can be seen as mutual approximations.

Strategies for approximate diagnosis can be used (1) to choose another, related notion of diagnosis when one definition of diagnosis fails (e.g. too many diagnoses, no diagnosis), (2) to reduce the cost of diagnosis using an anytime algorithm, (3) to deal with incompleteness of data and knowledge (4) to find an appropriate definition suited for the purpose and circumstance of performing diagnosis. See [van Harmelen and ten Teije, 1995] for more motivation for approximations in diagnosis.

In the literature, the definition of diagnosis is usually characterised using the logical entailment relation (Sec. 2 of this paper). In this paper we use existing notions of approximate entailment [Schaerf and Cadoli, 1995] (Sec. 3) to define notions of approximation in diagnosis (Sec. 4). In Sec. 5 we give four strategies for exploiting approximations in diagnosis using approximate entailment, and we illustrate these strategies in a small car domain example. The final section discusses related work and concludes.

2 Definition of Diagnosis

We use a common definition of diagnosis that is widespread in the literature. We follow [Console and Torasso, 1991] and combine in our definition both abductive and consistency based diagnosis, which accounts for a large variety of diagnostic systems from the literature.

Definition 1 (Diagnosis problem and solution)
Given a behaviour model BM (a logical theory in clausal form), and two sets of observations O^+ and O^- (both sets of literals read as conjunctions), a solution to a diagnostic problem is a set of literals E ("E" for explanation, again read as a conjunction), which satisfies the following:

$$\text{ABD}: \quad BM \cup E \vdash O^+ \tag{1}$$

$$\text{ABD}: \quad BM \cup E \nvdash \bot \tag{2}$$

$$\text{CBD}: \quad BM \cup E \cup O^- \nvdash \bot \tag{3}$$

We will write OBS *for the set of all possible observables from which the letters of O^+ and O^- must be taken, and require that E is disjoint from* OBS. *O^+ is the set of observations that must be explained abductively (i.e. they must be implied by the explanation E), while O^- only needs to be consistent with the explanation E.*

Formulae (1) and (2) constitute the abductive part of our notion of diagnosis (ABD), and (3) the consistency based part (CBD). Although (2) directly follows from (3) for classical entailment we include both conditions explicitly, because the central idea of our method of approximations in diagnosis is to parameterise the notion of diagnosis over different approximations of the entailment relation (in particular of Schaerf & Cadoli's approximate entailment relations).

We emphasise that our particular definition of a diagnostic problem and its solution is not of crucial importance to our *central* message that approximate entailment can be usefully exploited for diagnostic reasoning to obtain interesting and efficient results.

3 Summarising approximate entailment

In this section we will summarise the work in [Schaerf and Cadoli, 1995], which defines the approximate entailment

relations that we will exploit for our work on diagnoses. Schaerf and Cadoli define two approximations of classical entailment, named $\vdash_1$ and $\vdash_3$ which are either unsound but complete ($\vdash_1$) or sound but incomplete ($\vdash_3$). By analogy, they sometimes write $\vdash_2$ for classical entailment. Both of these approximations are parameterised over a set of predicate letters S (written $\vdash_1^S$ and $\vdash_3^S$) which determines their accuracy. We repeat some of the basic definitions from [Schaerf and Cadoli, 1995]:

Definition 2 (1-S-assignment, 3-S-assignment)
A 1-S- and 3-S-assignment are defined as follows:
- *If $x \in S$ then x and $\neg x$ get opposite truth values*
- *If $x \notin S$ then*
  - *for a 1-S-assignment, x and $\neg x$ both become 0.*
  - *for a 3-S-assignment, x and $\neg x$ do* not *both become 0.*

In other words: for letters in S, these assignments behave as classical truth assignments, while for letters $x \notin S$ they make either all literals false (1-S-assignments) or make one or both of x and $\neg x$ true (3-S-assignments).

Satisfaction of a clause by a 1-S- or 3-S-assignment, and the notions of 1-S-entailment and 3-S-entailment are defined in the same way as classical satisfaction and entailment.

Intuitively, for 3-S-entailment the predicates outside S are deemed irrelevant for deduction, while for 1-S-entailment these predicates are taken as false. The following syntactic notions can be used to clarify these definitions. For a theory in clausal form, 1-S-entailment corresponds to classical entailment, but after removing from every clause any literals with a letter outside S. When this results in an empty clause, the theory becomes the inconsistent theory $\bot$. Similarly, 3-S-entailment corresponds to classical entailment, but after removing every clause from the theory that contains a literal with a letter outside S. This may result in the empty theory $\top$.

The main result of [Schaerf and Cadoli, 1995] is:

Theorem 1 (Approximate entailment)

$$\vdash_3^{\emptyset} \Rightarrow \vdash_3^S \Rightarrow \vdash_3^{S'} \Rightarrow \vdash_2 \Rightarrow \vdash_1^{S'} \Rightarrow \vdash_1^S \Rightarrow \vdash_1^{\emptyset}$$

where $S \subseteq S'$. (Everywhere primed letters are a superset of the unprimed letters).

This states that $\vdash_3^S$ is a sound but incomplete approximation of the classical $\vdash_2$. The counterpositive of the second half of the theorem (reading $\not\vdash_1^S \Rightarrow \not\vdash_1^{S'} \Rightarrow \not\vdash_2$) states that $\not\vdash_1^S$ is a sound but incomplete approximation of $\not\vdash_2$.

Example 1 (Illustrating $\vdash_3^S$ and $\not\vdash_1^S$)
We illustrate these notions with figure 1. We can see that $\vdash_3^S$ is incomplete with respect to $\vdash_2$, since in the theory BM of figure 1 we have that classically $BM \cup \{H_3\} \vdash_2 O_1$, but if we restrict S to $\mathrm{LET}(BM) \setminus \{H_3\}$, where $\mathrm{LET}(BM)$ stands for all the predicate letters in BM, we do not have that $BM \cup \{H_3\} \vdash_3^S O_1$. This is so because taking $S = \mathrm{LET}(BM) \setminus \{H_3\}$ amounts to removing $H_3 \rightarrow O_1$ from BM.

$$\begin{array}{lllll} \neg H_1 & \vee & \neg H_0 & \vee & O_1 \\ \neg H_1 & \vee & \neg H_0 & \vee & O_2 \\ \neg H_2 & \vee & O_2 & & \\ \neg H_3 & \vee & O_1 & & \\ H_1 & \vee & \neg H_0 & \vee & O_3 \\ \neg H_4 & \vee & O_3 & & \\ \neg H_4 & \vee & \neg O_2 & & \end{array}$$

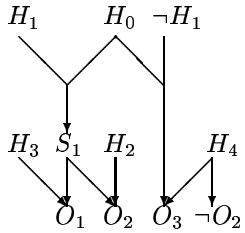

Figure 1: An example behaviour model (BM) formalised in clausal notation. In the formalisation of the network intermediate nodes (in this case only S_1) have been removed.

Similarly, $\not\vdash_1^S$ is incomplete with respect to $\not\vdash_2$ (or equivalently, $\vdash_1^S$ is unsound w.r.t. $\vdash_2$) since, for example, if $S = \mathrm{LET}(BM) \setminus \{H_0\}$, then $BM \cup \{H_1\} \not\vdash_2 O_1$, but $BM \cup \{H_1\} \vdash_1^S O_1$. This is so because taking $S = \mathrm{LET}(BM) \setminus \{H_0\}$ amounts to removing H_0 as a conjunct from $H_1 \wedge H_0 \rightarrow O_1$.

Furthermore, with increasing S, the accuracy of these approximations improves, until the approximate versions coincide with classical entailment when all letters are included in S.

Schaerf and Cadoli also give incremental algorithms for computing $\vdash_1^S$ and $\vdash_3^S$ when S increases. They have obtained attractive complexity results which state that even when computing $\vdash_2$ through iterative computation of $\vdash_3^S$, the total cost of the iterated computation is not larger than the direct computation of $\vdash_2$ (and similarly for $\not\vdash_1^S$ to compute $\not\vdash_2$). However, the iterative computation of the approximate entailment has as important advantage that the iteration may be stopped when a confirming answer has already obtained for a smaller value of S. This yields a potentially drastic reduction of the computational costs. The size of these savings depend on the appropriate choice for S.

Although the summary above is based on a propositional calculus, the theory that we will apply the approximations to in this paper is first-order (Fig. 3). In [Schaerf and Cadoli, 1995] they show how the propositional results can be extended to the first-order case in a straightforward way.

4 Summarising approximate entailment in diagnosis

In this section we summarize the results [ten Teije and van Harmelen, 1996] on applying $\vdash_1^S$ and $\vdash_3^S$ in diagnosis. We use $\vdash_1^S$ and $\vdash_3^S$ in both the ABD-part of our definition of diagnosis (written as ABD_1^S, ABD_3^S) and the CBD-part (written as CBD_1^S, CBD_3^S). Since we write $\vdash_2$ for the classical entailment relation, we will also write ABD_2 and CBD_2.

The main intuitions behind using $\vdash_1^S$ and $\vdash_3^S$ in diagnosis are as follows. By using $\vdash_1^S$, candidate solutions more easily satisfy part (1) of our definition of diagnosis, because $\vdash_2 \Rightarrow \vdash_1^S$. Similarly, by using $\vdash_3^S$, candidate solutions more

easily satisfy parts (2) and (3) of our definition of diagnosis, since $\not\vdash_2 \Rightarrow \not\vdash_3^S$.

We will write ABD_i^S when we intend both ABD_1^S and ABD_3^S, and similarly for CBD_i^S. Furthermore, we write $\mathcal{ABD}_i^S$ for the set of all diagnoses E which satisfy $\text{ABD}_i^S(E)$, and similarly for $\mathcal{ABD}_2$, $\mathcal{CBD}_i^S$ and $\mathcal{CBD}_2$

There are two important relations $\Subset$ and $\Supseteq$ for relating the ABD_i^S and CBD_i^S diagnoses.

Definition 3 *For any set of sets P and P', $P \Subset P'$ and $P \Supseteq P'$ are defined by:*

$$P \Subset P' \equiv \forall p \in P\ \exists p' \in P' : p \subseteq p'$$
$$P \Supseteq P' \equiv \forall p' \in P'\ \exists p \in P : p \subseteq p'$$

Notice that these relations are relations between sets of sets. The required superset and subset relation is among the elements (sets) of these sets of sets.

Example 2 (Examples of $\Subset$ and $\Supseteq$)

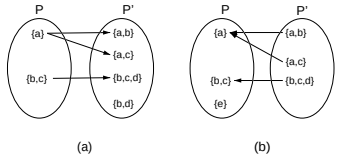

Figure (a): $P \Subset P'$, figure (b): $P \Supseteq P'$.

For the set of abductive diagnoses we have the following relation:

Theorem 2 (Relations between $\mathcal{ABD}_i^S$)

$$\begin{aligned}\emptyset = \ & \mathcal{ABD}_1^\emptyset \subseteq \mathcal{ABD}_1^S \Subset \mathcal{ABD}_1^{S'} \Subset \mathcal{ABD}_2 \\ & \mathcal{ABD}_2 \Supseteq \mathcal{ABD}_3^{S'} \Supseteq \mathcal{ABD}_3^S \Supseteq \mathcal{ABD}_3^\emptyset = \emptyset\end{aligned}$$

This states that ABD_1^S diagnoses consist of parts of ABD_2 diagnoses, and that ABD_3^S diagnoses contain ABD_2 diagnoses. Another result is on the number of diagnoses:

Theorem 3 (Sizes of $\mathcal{ABD}_i^S$)

$$\begin{aligned}0 = \ & |\mathcal{ABD}_1^\emptyset| \leq |\mathcal{ABD}_1^S| \leq |\mathcal{ABD}_1^{S'}| \leq |\mathcal{ABD}_2| \\ & |\mathcal{ABD}_2| \geq |\mathcal{ABD}_3^{S'}| \geq |\mathcal{ABD}_3^S| \geq |\mathcal{ABD}_3^\emptyset| = 0\end{aligned}$$

We have analogous results for the CBD-part:

Theorem 4 (Relations between CBD_i^S)

$$\begin{aligned}&\emptyset = \mathcal{CBD}_1^\emptyset \subset \mathcal{CBD}_1^S \Subset \mathcal{CBD}_2 \Subset \mathcal{CBD}_3^S \Subset \mathcal{CBD}_3^\emptyset = \mathcal{E} \\ &\emptyset = \mathcal{CBD}_1^\emptyset \subset \mathcal{CBD}_1^S \subset \mathcal{CBD}_2 \subset \mathcal{CBD}_3^S \subset \mathcal{CBD}_3^\emptyset = \mathcal{E} \\ &0 = |\mathcal{CBD}_1^\emptyset| < |\mathcal{CBD}_1^S| < |\mathcal{CBD}_2| < |\mathcal{CBD}_3^S| < |\mathcal{CBD}_3^\emptyset|\end{aligned}$$

where $\mathcal{E}$ stands for any consistent set of literals whose letters are taken from $\text{LET}(BM) \setminus \text{OBS}$. The first sequence of inclusions states that CBD_1^S diagnoses consist of parts of classical CBD-diagnoses, and that every classical diagnosis is contained in at least one CBD_3^S diagnosis. In [ten Teije and van Harmelen, 1996] we also have theorems about the type (superset or subset) of new diagnoses that can be found by changing S. *"New"* means that using the new value of S, we compute at least one superset/subset diagnosis which was not present for the old value of S. These theorems are summarized in Fig. 2.

diagnosis definition	change of S	new superset diagnosis	new subset diagnosis	nr.
ABD_1^S	$S \rightarrow S'$	yes	no	more
$\text{ABD}_1^{S'}$	$S' \rightarrow S$	no	only	less
ABD_3^S	$S \rightarrow S'$	no	yes	more
$\text{ABD}_3^{S'}$	$S' \rightarrow S$	only	no	less
CBD_1^S	$S \rightarrow S'$	only	no	more
$\text{CBD}_1^{S'}$	$S' \rightarrow S$	no	no	less
CBD_3^S	$S \rightarrow S'$	no	no	less
$\text{CBD}_3^{S'}$	$S' \rightarrow S$	no	only	more

Figure 2: Summarising some results of using approximate entailment in the diagnosis definition [ten Teije and van Harmelen, 1996]. *"yes"* means that using the new S results in superset/subset diagnoses, and similarly for *"no"*. *"only"* means that all the new computed diagnoses are superset/subset diagnoses. *more* and *less* means that the number of diagnoses increases and decreases respectively.

5 Strategies for approximate diagnosis

5.1 General strategies

We can use approximate diagnosis (results of Sec. 3 and 4) for solving problems of too many, too few, too large and too small diagnoses. When such a problem occurs with a particular notion of diagnosis, we could choose another related notion. In our case of using Schaerf & Cadoli's approximate entailment relation, this means changing the parameter S.

In this section we only consider increasing S, because this allows us to use the incremental algorithm of Schaerf & Cadoli. They show that the total cost of the iterated computation is not larger than directly computing classical entailment. However, the iterative computation of the approximate entailment may be stopped when a satisfactory answer is already obtained for a smaller value of S.

Focusing on increasing S results in the following general strategies:

Solutions for the problem of too few diagnoses:
- Shifting from ABD_1^S to $\text{ABD}_1^{S'}$
- Shifting from ABD_3^S to $\text{ABD}_3^{S'}$ (see strategy III)
- Shifting from CBD_1^S to $\text{CBD}_1^{S'}$ (see strategy II).

Solutions for the problem of too many diagnoses
- Shifting from CBD_3^S to $\text{CBD}_3^{S'}$ (see strategy I)

Solutions for the problem of too small diagnoses
- Shifting from ABD_1^S to $\text{ABD}_1^{S'}$ (see strategy IV)
- Shifting from CBD_1^S to $\text{CBD}_1^{S'}$

Solution for the problem of too large diagnoses

• Shifting from ABD_3^S to $\text{ABD}_3^{S'}$

These strategies are general in the sense that they do not exploit specific properties of the behaviour model. Using such properties would be more attractive because this enables us to be more precise about how we could extend S and determine the characteristics of the diagnoses that will be computed. In Sec. 5.3 we give such specific strategies.

5.2 Example Behaviour Model (BM)

The example that we use for our strategies is taken from [Dupré, 1994] and is shown in Fig. 3. This figure shows a partial causal model of a car. The causal network contains 42 nodes and 40 causal links. We transform the causal network of Fig. 3 to an equivalent two layered network, because we use the results of [ten Teije and van Harmelen, 1996], and some of them are restricted to a two layered network.

In causal networks [Console and Torasso, 1990], states are represented as predicates, and the fact that state S_i necessarily causes state S_j is represented as $S_i \rightarrow S_j$. The fact that S_i *possibly* causes state S_j can be written as $S_i \wedge \alpha_{ij} \rightarrow S_j$, where α_{ij} (called the incompleteness assumption) is interpreted as the unknown condition required for S_i to cause S_j.

The letters of the two layered version of the network from Fig. 3 are the initial causes (written as the set $\mathcal{H}$), the incompleteness assumptions (the set $\mathcal{A}$), and the observables (the set $\mathcal{O}$).

5.3 Strategies Dependent on properties of BM

In this section we give four strategies of approximations in diagnosis which depend on properties of the behaviour model.

Strategy (I) and (II) use the specificity of observables, strategy (III) and (IV) use the necessity of the causal relations. Besides using different properties of the behaviour model, the strategies are distinct examples of approximating diagnoses. Strategy (I) and (II) are examples of changing the CBD-part. Strategy (III) and (IV) are examples of changing the ABD-part. Also, they are strategies to deal with different types of problems with the diagnoses: strategy (I) reduces the number of diagnoses, strategy (II) and (IV) increase the number of diagnoses, strategy (III) increases the size of the diagnoses.

Using Specificity of Observations

Strategy (I) and (II) are based on the specificity of observables. We use a whole spectrum from specific to a-specific observables. We call an observable more specific if it has fewer possible causes. For the car example this spectrum is shown in Fig. 4.

Strategy (I). This strategy can solve the problem of "too many diagnoses". The strategy is based on the idea that an explanation cannot be a diagnosis if one of its specific observables is not observed. We approximate this by using the spectrum from the specific to the a-specific end. We apply definition 1 using CBD_3^S, ABD_2, and the initial value of S (S_{init}) is only all the possible causes and the incompleteness assumptions, and none of the observables. Because the consistency part of the ABD-formula (2) implies $\text{CBD}_3^{S_{init}}$, any ABD-explanation will be a diagnosis. Extending S with the most specific observables means that a cause of such an observable can only remain part of the diagnosis when its specific observables are observed. We extend S with increasingly less specific observables until one diagnosis is left or no diagnosis is left. In the latter case we consider the diagnoses of the previous step as the best diagnoses.

Example 3 (Strategy (I)) *(In the examples we give only the subset minimal diagnoses.) We illustrate this strategy using the following values:*

$$
\begin{array}{ll}
O^+ = & \{\textit{oil warning light(red)}\} \\
O^- = & \{\ \neg\textit{exhaust smoke(black)}, \\
 & \quad \neg\textit{accelerator response(delayed)}\} \\
S_{init} = & \mathcal{H} \cup \mathcal{A}
\end{array}
$$

S	$Diagnoses$
S_{init}	{*engine mileage(*$>$ *100000km)*,α_2},
	{*engine mileage(*$>$ *100000km)*,α_3},
	{*road cond(poor), gr. clearance(low)*,α_1}
$+Spec_1$	{*engine mileage(*$>$ *100000km)*,α_2},
	{*engine mileage(*$>$ *100000km)*,α_3},
	{*road cond(poor), gr. clearance(low)*,α_1}
$+Spec_2$	{*road cond(poor), gr. clearance(low)*,α_1}

For S_{init} each ABD*-explanation is* CBD_3^S*-consistent with O^-, because S contains no observables. We do not introduce inconsistency by extending S with $Spec_1$, in contrast with extending S further with $Spec_2$. This extension introduces inconsistency between* ¬exhaust smoke(black) *and each of the explanations* {engine mileage($>$ 100000km),α_2} *and* {engine mileage($>$ 100000km),α_3}. *We stop at $S_{init} + Spec_1 + Spec_2$ because we have just one diagnosis left.*

Our final explanation is CBD_3^S*-consistent with the negative observation* ¬ accelerator response(delayed) *(for $S = S_{init} + Spec_1 + Spec_2$). Under* CBD_2 *it would have been inconsistent with this observation, but we consider this is of less importance because of the low specificity of this observable.*

Strategy (II). This strategy can solve the problem of "too few diagnoses". The strategy is based on the idea that a cause has to be excluded if a specific observable is not observed. We apply definition 1 using CBD_1^S and ABD_2. If an observable is not in S, then no possible cause of this observable can be part of the diagnosis. We allow non-observed observables to be implied by the diagnosis only if these additional observables are non-specific. Therefore the strategy uses the spectrum of specificity of observables from a-specific to specific.

We initialise S with all possible causes, plus $O^+$and O^-. Excluding O^+ from S would result in inconsistency of all the explanations, because each possible cause of an O^+ observable would be excluded by CBD_1^S. Excluding O^- would

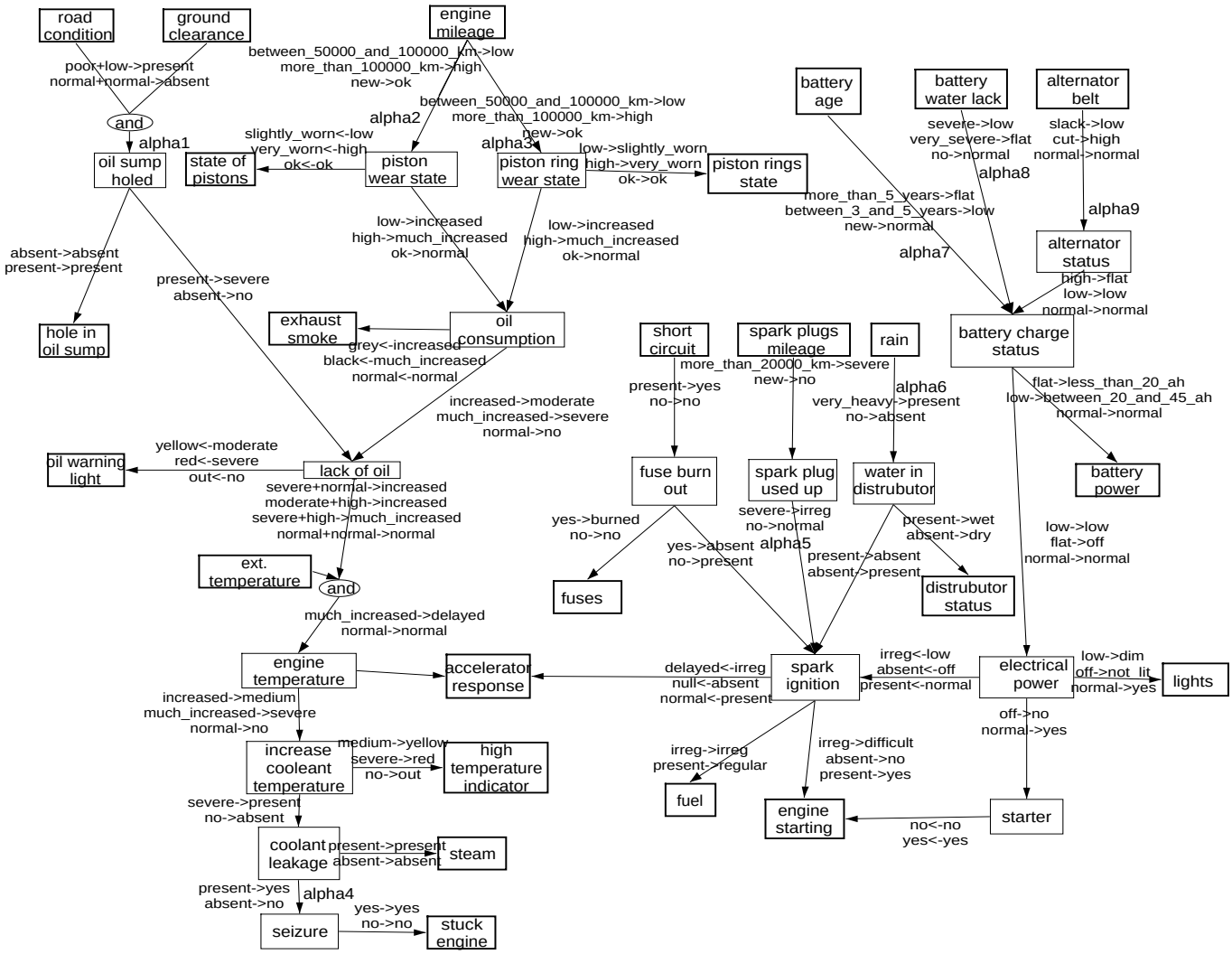

Figure 3: Behaviour model of a car from [Dupré, 1994]. The bold lined boxes are initial causes and observables. For example, the top-right-most causal link corresponds to the formulae: $alternatorbelt(slack) \wedge \alpha_9 \rightarrow alternatorstatus(low)$, $alternatorbelt(cut) \wedge \alpha_9 \rightarrow alternatorstatus(high)$ $alternatorbelt(normal) \wedge \alpha_9 \rightarrow alternatorstatus(normal)$

mean inconsistency with O^- and BM. We increasingly extend S with more specific observables. This introduces possibly more diagnoses, because more causes will be enabled by the observables in S.

Example 4 (Strategy (II)) *We illustrate this strategy using the following values:*

$$
\begin{aligned}
O^+ &= \{\textit{fuel(irreg)}\} \\
O^- &= \{\neg\textit{fuses(burned)}\} \\
S_{init} &= \mathcal{H} \cup \mathcal{A} \cup O^+ \cup O^-
\end{aligned}
$$

S	$Diagnoses$
S_{init}	$none$
$+Spec_5$	$none$
$+Spec_4$	{*spark plugs mileage(*$>$ *200000km)*,α_5}
$+Spec_3$	{*spark plugs mileage(*$>$ *200000km)*,α_5}, {*battery age(between 3 and 5 years)*,α_7}, {*battery water lack(severe)*,α_8}, {*alternator belt(slack)*,α_9}

S_{init} gives no diagnoses because every cause still has an observable which is not in S_{init}. S_{init} extended with $Spec_5$ and $Spec_4$ results in the first diagnosis. All observables of the causes spark plugs mileage *and α_5 are in S. If we continue extending S more diagnoses could be found. This example illustrates that one would prefer the diagnosis whose observables are observed (consistent) or (if they are not observed) not very specific.*

most specific 1 poss. cause ($Spec_1$)	2 poss. causes ($Spec_2$)	3 poss. causes ($Spec_3$)	6 poss. causes ($Spec_4$)	**least specific** 9 poss. causes ($Spec_5$)
hole in oil sump state of pistons piston ring state fuses distributor status	exhaust smoke	oil warning light battery power lights high temp.indic. steam stuck engine	fuel engine starting	accel. response

Figure 4: Specificity of the observables in the car example. This spectrum is from specific observables (1 possible cause) to a-specific observables (9 possible causes).

Using Necessity of Causal Relations

Strategies (III) and (IV) are based on the necessity of the causal relation.

Our behaviour model contains necessary causal relations ($S_i \rightarrow S_j$) and possible causal relations ($S_i \wedge \alpha_{ij} \rightarrow S_j$). However, for our strategies we will use a whole spectrum of the necessity of the causal relation by dividing the incompleteness assumptions α_i in several groups and ordering them. This ordering is meant to indicate the degree of necessity of the causal relation. This ordering in the car domain is as follows:

Necessary ($Poss_0$)	($Poss_1$)	($Poss_2$)	**Possible** ($Poss_3$)
No α_i in in causal link	α_5 α_9	α_2,α_3,α_4 α_7,α_8	α_1 α_6

Note that this spectrum is domain specific knowledge, whereas the specificity of observables can be determined syntactically.

Strategy (III). This strategy can solve the problem of "too few diagnoses". The strategy is based on the idea of taking as little notice as possible of non-necessary relations. We use the spectrum from the necessary-side to the possible-side. This amounts to first completely ignoring the possible relations and introducing them increasingly starting from the most necessary ones. We apply definition 1 using ABD_3^S, CBD_2, and the initial value of S is all symbols of BM without the α_i. This results in only diagnoses which use necessary causal relations. If we extend S with α_i, we also use less necessary causal relations.

Example 5 (Strategy (III)) *We illustrate this strategy using the following values:*

$$\begin{aligned} O^+ &= \{\textit{battery power(between 20 and 45 ah)}\} \\ O^- &= \emptyset \\ S_{init} &= \mathcal{H} \cup \mathcal{O} \end{aligned}$$

S	$Diagnoses$
S_{init}	$none$
$+Poss_1$	{*alternator belt(slack)*,α_9}
$+Poss_2$	{*battery age(between 3 and 5 years)*,α_7}, {*battery water lack(severe)*,α_8}, {*alternator belt(slack)*,α_9}

In the first step only necessary relations are used, and no diagnosis is found. Extending S with the incompleteness assumptions with the highest degree of necessity ($Poss_1$) enables the use of the causal relations with α_5 and α_9 giving the diagnosis {alternator belt(slack),α_9}. *Adding more incompleteness assumptions allows us to use more possible relations and results in two extra diagnoses.*

Strategy (IV). This strategy can solve the problem of "too small diagnoses". The strategy is based on the idea to start with paying no attention to the necessity of relations, i.e. to use all relations as necessary. All diagnoses will be without the incompleteness assumptions even though they might be using non-necessary links. These diagnoses can later be extended using the necessity of the causal relations. We apply definition 1 using ABD_1^S, CBD_2, and the initial value of S is all letters of BM excluding the α_i. Extending S with those α_i which are in the spectrum at the *possible* end, results in detailed diagnoses if such relations are used in the explanation. This means that those diagnoses with the most unreliable causal links will be the first to get extended with the appropriate α_i.

We do not illustrate this strategy because of lack of space.

6 Discussion

The main message of this paper is that we apply approximation strategies to diagnosis, and that the approximation strategies are informed by particular properties of the domain knowledge. The strategies (I)–(IV) all deal with problems concerning the size and number of the diagnoses. However, our approximation techniques can also be used to model many of the general focusing strategies described in the literature. For instance, by dividing the behaviour model in submodels, and by choosing for S only the letters of a particular sub-model, we effectively obtain a focusing strategy based on

the use of these sub-models. Other focusing strategies can be dealt with in the same way.

A first obvious task for future work would be to apply our proposed algorithms to a more realistic application. A candidate for this could be a domain where explanations can be ordered based on their urgency. According to theorem 2, using ABD_1^S and initialising S with the most urgent explanation candidates, and adding gradually less urgent candidates computes only urgent subsets of classical diagnosis, and computes only non-urgent abductive diagnoses when S is increased if resources allows. This yields an anytime algorithm that performs well for urgent diagnoses under time pressure. A second obvious task is to study the efficiency behaviour of our approximation algorithms in larger behaviour models then we have presented here.

Finally, our approach to diagnostic strategies should be compared with other approaches, in particular [Böttcher and Dressler, 1994], based on [Struss, 1992]. In this approach, so called "working hypotheses" indicate various restrictions on, or preferences for potential diagnoses. The set of active working hypotheses is then modified to switch from one set of diagnoses to another. We claim that defining S as the set of all letters from the behaviour model plus the active working hypotheses would yield an alternative formalisation.

References

[Böttcher and Dressler, 1994] C. Böttcher and O. Dressler. A framework for controlling model-based diagnosis systems with multiple actions. *Annals of Mathematics and Artificial Intelligence*, 11(1-4), 1994.

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

[Console and Torasso, 1991] L. Console and P. Torasso. A spectrum of logical definitions of model-based diagnosis. *Computational Intelligence*, 7(3):133–141, 1991.

[Dupré, 1994] D.Theseider Dupré. *Characterizing and Mechanizing Abductive Reasoning*. PhD thesis, Universita di Torino, 1994.

[Schaerf and Cadoli, 1995] M. Schaerf and M. Cadoli. Tractable reasoning via approximation. *Artificial Intelligence*, 74(2):249–310, April 1995.

[Struss, 1992] P. Struss. What's in SD? Towards a theory of modeling for diagnosis. In L. Console, J.H. de Kleer, and W.C. Hamscher, editors, *Readings in Model-based Diagnosis*, pages 419–449. Morgan Kaufmann, 1992.

[ten Teije and van Harmelen, 1996] A. ten Teije and F. van Harmelen. Computing approximate diagnoses by using approximate entailment. In *Proc. of the Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'96)*, Boston, Massachusetts, November 1996.

[van Harmelen and ten Teije, 1995] F. van Harmelen and A. ten Teije. Approximations in diagnosis: motivations and techniques. In A. Levy and P. Nayak, editors, *Proc. of SARA-95, Symposium on Abstraction, Reformulation, and Approximation*, pages 149–155, Quebec, Aug 1995.