Proceedings of the Symposium on Abstraction, Reformulation and Approximation,
(SARA'95), Ville d'Esterl, Canada, August 1995

Approximations in diagnosis: motivations and techniques

Frank van Harmelen
SWI
University of Amsterdam
Roetersstraat 15
1018 WB Amsterdam
frankh@swi.psy.uva.nl

Annette ten Teije
SWI
University of Amsterdam
Roetersstraat 15
1018 WB Amsterdam
annette@swi.psy.uva.nl

Abstract

We argue that diagnosis should not be seen as solving a problem with a unique definition, but rather that there exists a whole space of reasonable notions of diagnosis. These notions can be seen as mutual approximations. We present a number of reasons for choosing among different notions of diagnosis. We also present an exhaustive categorisation of techniques that can be employed to obtain approximations, as well as a number of specific example techniques for each category. We also show that it is possible to characterise the relations between the approximations obtained by these techniques.

1 Introduction

The AI literature contains many definitions for diagnostic reasoning. Implicitly or explicitly, these papers make many assumptions about the kind of knowledge that is assumed to be available, and the kind of properties that a diagnosis is supposed to satisfy. Tacitly these papers claim that if these assumptions are satisfied, they present *the* appropriate definition of diagnosis. However, there are two 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. Firstly, the choice of an appropriate definition of diagnosis depends on the purpose for which diagnosis is performed, and the circumstances under which this must be done. Secondly, both data and knowledge involved in diagnosis are often incomplete. The resulting failure of one definition of diagnosis can often be dealt with by choosing another, related notion of diagnosis.

In this paper, we view different definitions of diagnosis as approximations of each other. This enables us to give (i) a number of concrete reasons for switching between different definitions, (ii) a categorisation of techniques to obtain approximations, and (iii) concrete techniques for each of these different approximation categories.

The structure of this paper is as follows. Sections 2 and 3 are introductory: Sec. 2 elaborates on the motivation for studying approximations in diagnosis and gives some initial examples, Sec. 3 recalls some standard definitions and introduces some terminology. The main substance of the paper is in Sec. 4–5. Section 4 lists a number of reasons which may cause us to move from any given definition of diagnosis to an approximation. Sec. 5 distinguishes categories of techniques for obtaining such approximations and gives extensive examples of each category. Section 6 discusses related work and Sec. 7 concludes.

2 Motivation

In this section we elaborate on the reasons which make approximations in diagnosis an important subject of study, and we give some examples of the use of such approximations.

The definition of diagnosis depends on purpose and circumstances. Diagnosis is often taken as "finding the reasons that explain observed abnormal behaviour". However, diagnosis is never performed for its own sake, and always serves some exterior motive like repair, damage prevention, symptom suppression, etc. Therefore, how the informal notion of diagnosis should be interpreted is strongly determined by the purpose for which we perform diagnosis, and the circumstances under which this must be done. We give three examples.

[PNdK94] has studied the behaviour of telephone operators at an emergency switchboard who must decide whether an emergency call requires an ambulance to be sent. These operators engage in a dialogue that leads to a diagnostic process, but their goal is not to perform a precise diagnosis. Given the dramatic consequences of erroneously failing to send an ambulance, their diagnostic reasoning stops as soon as the set of candidate hypotheses contains a single cause which merits an ambulance. Thus, they approximate a precise diagnosis by checking only if an urgent cause is a possibility.

The converse happens when a physician has to decide whether or not to administer a severe drug treatment with possibly dangerous side effects. Such a high-risk decision is only taken if all the other diagnoses that might lead to less severe treatments have been ruled out. Thus, here diagnosis is interpreted as reasoning until only one conclusion remains, quite the converse of the behaviour of the switch board operators.

A third and final example can be found in the behaviour of a computer technician who diagnosis a faulty computer. In order to localise the fault, technicians may reason in terms of individual components, and will entertain different hypotheses consisting of sets of faulty components. However, as soon as all remaining competing hypotheses contain components from the same physical board, technicians will abort their diagnostic reasoning and simply replace the board. Here again, we see how the purpose of the diagnosis (in this case repairing the computer by replacing a board) causes the technician to be satisfied with an approximate diagnosis instead of a precise diagnosis.

Incompleteness of data and knowledge. In an ideal world, we would have complete knowledge about the object under diagnosis and its observed behaviour. We would then be able to uniquely identify the set of causes responsible for this behaviour. However, in practice both knowledge and observations are incomplete. As a result, diagnosis often yields multiple candidate causes among which we cannot further distinguish, or, perhaps worse, it may fail to yield any candidate at all. In both cases, we must consider approximations of such results in order to enable further action. We give an example of either case.

Consider a doctor who cannot further discriminate among competing sets of disorders that are responsible for a set of symptoms. Unfortunately, each candidate set would require a different treatment and applying all treatments simultaneously is impossible because of interactions among the drugs involved. A reasonable approximation this doctor may choose is to ignore some of the less severe symptoms, and see if the remaining set of severe symptoms is amenable to a single treatment. Similarly, when no single explanation can be found, reducing the set of observations may enable an explanation of at least the most urgent subset of the symptoms.

The above has illustrated how both the purpose of diagnosis and the circumstances under which it is performed, as well as the incompleteness of both data and knowledge may cause us to consider alternative definitions of diagnosis which can be seen as mutual approximations.

3 Basic Definitions

Here we introduce a number of standard definitions from the diagnostic literature. Our definitions are based on [CT90]. We assume that the model of the system under diagnosis (called the system description, or SD) is defined as a causal network plus a definition of diagnostic labels:

System description: A system description SD is a set of formulae SD $=$ CN $\cup$ DL, where CN is a causal network and DL is a definition of diagnostic labels.

Causal Network: Let $\mathcal{S} = \{S_1, S_2, \ldots\}$ be a set of predicate letters called *causal states*, let $\mathcal{C} = \{C_1, C_2, \ldots\}$ be a set of predicate letters called *condition symbols*, to describe observations which need not be explained, and let $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots\}$ be a set of predicate letters called *assumption symbols* that represent unknown conditions, then a *causal network* CN is a set of implications of one of the following two forms:

$$\bigwedge X_i \quad \rightarrow \quad S_k \tag{1}$$

$$(\bigwedge X_i) \wedge \alpha_j \quad \rightarrow \quad S_k. \tag{2}$$

with the X_i taken from $\mathcal{S} \cup \mathcal{C}$ (and at least on X_i from $\mathcal{S}$). Implications of form (1) represent *necessary causality*, and those of form (2) represent *possible causality*. If in a clause all X_i are taken from $\mathcal{S}$ (i.e. no condition symbols are present), we speak of *unconditional causality*, and of *conditional causality* otherwise. Intuitively, a clause of the form $C_1 \wedge S_1 \wedge S_2 \wedge \alpha_1 \rightarrow S_3$ is to be read as: under condition C_1, states S_1 and S_2 together may cause state S_3 (depending on the unknown truth of α_1). If α_1 were not present, $S_1 \wedge S_2$ would necessarily cause S_3 whenever C_1 were true.

Diagnostic labels: Let $\mathcal{L} = \{L_1, L_2, \ldots\}$ be a set of predicate letters called *diagnostic labels*, then a *diagnostic label definition* is a set of equivalences of the form $\bigwedge_i S_i \leftrightarrow L_k$. We will write *def*$(L_k)$ for the set of the S_i. A diagnostic label defines a name for a set of states that somehow "belong together" because they form a known syndrome, or because they require the same treatment or any other reason.

Coherent: A set of states STATES $\subseteq \mathcal{S}$ is called coherent *iff* there exists a set of labels LBLS $\subseteq \mathcal{L}$ such that DL $\vdash$ STATES $\leftrightarrow$ LBLS[1], in other words: a coherent set of states can be exactly characterised by a set of diagnostic labels.

Observations: The set of *observations* $\mathcal{O}$ of a causal network CN are those S_i that only occur in the right-hand side of implications in CN. These are the states that must be explained.

These different categories of knowledge can then be used to give particular definitions of a diagnostic problem and its solution.

[1] A set of formulae that occurs in a formula is to be read as a conjunction.

Diagnostic problem: A diagnostic problem DP is a tuple $\langle \mathrm{SD}, \mathrm{OBS}, \mathrm{CXT}\rangle$, with SD a system description, $\mathrm{OBS} \subseteq \mathcal{O}$ and $\mathrm{CXT} \subseteq \mathcal{C}$.

Diagnostic Relation and solution:. A triple $\mathrm{SOL} = \langle \mathrm{ASS}, \mathrm{STATES}, \mathrm{LBLS}\rangle$ with $\mathrm{ASS} \subseteq \mathcal{A}$, $\mathrm{STATES} \subseteq \mathcal{S}$ and $\mathrm{LBLS} \subseteq \mathcal{L}$ is a solution to a diagnostic problem DP under diagnostic relation DR *iff* relation DR holds between DP and SOL, written $\mathrm{DR}(\mathrm{DP}, \mathrm{SOL})$. A particular diagnostic relation DR may only involve one or two of the components of SOL, in which case we shall write SOL as a unary or binary tuple.

Example: Abductive relation and solution. A common example of a diagnostic relation and solution is the standard notion of abduction from e.g. [CT90] in which a solution is a set of assumptions symbols and diagnostic labels which imply all of the observed observations, and do not imply any of the absent observations. $\mathrm{SOL} = \langle \mathrm{ASS}, \mathrm{LBLS}\rangle$ is a solution for diagnostic problem $\mathrm{DP} = \langle\langle \mathrm{CN}, \mathrm{DL}\rangle, \mathrm{OBS}, \mathrm{CXT}\rangle$ under the abductive diagnostic relation *iff*:

$$\mathrm{CN} \cup \mathrm{DL} \cup \mathrm{CXT} \cup \mathrm{ASS} \cup \mathrm{LBLS} \vdash \mathrm{OBS}^{+} \text{ and} \quad (3)$$

$$\mathrm{OBS}^{-} \cup \mathrm{CN} \cup \mathrm{DL} \cup \mathrm{CXT} \cup \mathrm{ASS} \cup \mathrm{LBLS} \nvdash \bot \quad (4)$$

where $\mathrm{OBS}^{+} = \mathrm{OBS}$ and $\mathrm{OBS}^{-} = \{\neg O | O \in \mathcal{O} \backslash \mathrm{OBS}\}$.

It is important to emphasise that although we assume that SD is a set of Horn clauses plus diagnostic labels, we do *not* assume any particular definition of a diagnostic relation DR. Indeed, DR is one of the concepts that we will vary in our study of approximations. The above definition of an abductive solution is only an *example* of a particular notion of diagnosis.

Finally, we introduce some terminology to speak about approximations:

Exact diagnosis: We call a solution SOL an *exact diagnosis* for a diagnostic problem DP under a diagnostic relation DR *iff* $\mathrm{DR}(\mathrm{DP}, \mathrm{SOL})$ holds.

Approximate diagnosis: We call SOL an *approximate diagnosis* for DP under DR *iff* $\mathrm{DR}'(\mathrm{DP}, \mathrm{SOL})$ holds for some approximation DR' of DR, or if $\mathrm{DR}(\mathrm{DP}', \mathrm{SOL})$ holds for some approximation DP' of DP. Thus, an approximate diagnosis for a given problem and diagnostic relation is an exact diagnosis for an approximation of the problem or of the relation. What approximations of diagnostic problems and relations are will be the topic of subsequent sections.

Over- and under-diagnosis: A diagnosis is an *over-diagnosis* (*under-diagnosis*)for a given problem DP if it accounts for a superset (subset) of the observations in DP. Of course, an approximate diagnosis may be simultaneously an over and an under diagnosis.

4 Reasons for computing approximate diagnoses

In this section, we enumerate a number of reasons for computing approximate diagnoses instead of only exact solutions. Whereas the motivating examples from Sec. 2 were all phrased in terms of application specific properties (e.g. the risk of not sending an ambulance), here we will phrase the reasons for computing approximate diagnoses in terms of domain independent properties of a diagnosis.

(R1) Reduce the number of diagnoses. In general, for a given DP and DR, the relation $\mathrm{DR}(\mathrm{DP}, \mathrm{SOL})$ may hold for more than one value of SOL. This is problematic if these competing values for SOL imply different actions (e.g. repair) to be taken. We can try to adjust DR to DR' so that $\mathrm{DR}'(\mathrm{DP}, \mathrm{SOL}')$ holds for fewer values SOL' (and similarly by adjusting DP). The solutions SOL$'$ can then be taken as approximations of the solutions SOL. The most commonly found example of this in diagnostic reasoning is known as Occam's Razor: among a number of competing diagnoses with the same explanatory power, select the simplest under some appropriate definition of "simple". In our terminology, this amounts to obtaining DR$'$ by adding the selection criterion to DR.

(R2) Increase the number of diagnoses. The converse of (R1) may also happen: $\mathrm{DR}(\mathrm{DP}, \mathrm{SOL})$ may not hold for any value SOL. A common example of this is that all combinations of causes fail to explain a symptom that has nevertheless been observed. A total lack of solutions would imply a total inability to act, eg. an inability to decide which treatment to administer to a patient. It is therefore quite common for a physician to accept a diagnosis even when not all symptoms have been accounted for. In our terminology, approximating DP by DP$'$ (by ignoring some observations) may lead to a solution $\mathrm{DR}(\mathrm{DP}', \mathrm{SOL}')$ whereas no solution existed for $\mathrm{DR}(\mathrm{DP}, \mathrm{SOL})$. Notice that in this case, SOL' is an under-diagnosis for DP.

(R3) Reduce the size of a diagnosis. Besides the number of alternative values of SOL for which $\mathrm{DR}(\mathrm{DP}, \mathrm{SOL})$ holds, further problems may be caused by the size of an individual solution SOL. This corresponds to the dilemma of a doctor whose patient suffers from a large number of diseases. When treating all diseases simultaneously is not possible, the doctor may choose to diagnose only the most severe complaints (ie. change from DP to DP$'$ with a reduced observation set), and solve $\mathrm{DR}(\mathrm{DP}', \mathrm{SOL}')$ instead, and treat only those diseases found in SOL$'$ (where presumably $\mathrm{SOL}' \subset \mathrm{SOL}$). Again, SOL' would be an under-diagnosis for DP.

(R4) Make a diagnosis coherent. The reason for introducing the diagnostic labels $\mathcal{L}$ in our system description (following [CT90]) is that in realistic applications, even when reasoning in terms of causal states is possible, meaningful diagnoses are often not phrased in terms of such states. Instead, a separate vocabulary is created to express the diagnostic solutions. Physicians often speak in terms of syndromes, which consist of causal states that together form a coherent disease pattern. Such syndromes are used in communication among physicians, and in therapy planning. In many domains, an incoherent diagnosis that does not correspond to a set of diagnostic labels is of as little value as no diagnosis at all, since no meaningful action can be associated with such an incoherent diagnosis. It is then attractive to try to find an approximate diagnosis which, although not exact, is at least coherent, and therefore allows us to undertake an action.

(R5) Make a diagnosis complete. We call a diagnosis *complete* if it contains at least all the real causes of the observed complaints (and perhaps some more). In practice, because our knowledge is often insufficient, we can of course seldom be certain that a diagnosis is complete. Nevertheless, in many domains it is of the utmost importance that a diagnosis is complete, because missing out some of the real causes may have dramatic consequences (eg. not sending an ambulance, in Sec. 2). It is then attractive to calculate approximations which may be too large, but which give us a higher confidence that the real causes of the observed behaviour are included in the approximation (perhaps at the expense of including additional irrelevant causes). In general, such approximations will be over-diagnoses.

(R6) Make a diagnosis sound. The converse of (R5) is to make a diagnosis sound (ie. to ensure that it contains only real causes, but perhaps not all of them). This is useful in applications where administering the wrong treatment (or performing the wrong repair) is as disastrous as taking no action at all. In such a case, we may want to compute approximate solutions which are so small that they perhaps do not include all causes, but of the causes they do include, we can be certain that they are among the real causes of the complaints. Diagnoses computed for this reason will often be under-diagnoses.

This concludes our list of reasons why approximate diagnoses may be attractive, and sometimes even more attractive than "exact" solutions.

5 Techniques for computing approximate diagnoses

We now proceed to describe a number of techniques that can be used for computing an approximate diagnosis. Our definition of an approximate diagnosis SOL' for a diagnostic problem DP and diagnostic relation DR was that either DR'(DP,SOL') or DR(DP',SOL') should hold for some approximation DP' of DP or DR' of DR. This enables us to categorise all the techniques for computing approximate diagnoses into two categories: those that vary DR and those that vary DP. Those that vary the diagnostic relation DR can be divided into three subclasses, depending on whether they adjust (a) the provability relation used in DR, (b) the definitions of the different categories of observations that occur in this relation, or (c) the selection of final solutions from among the possible candidate solutions of this relation. The other techniques wich vary the diagnostic problem DP can be subdivided on the basis of the part of $\mathrm{DP} = \langle \mathrm{SD}, \mathrm{OBS}, \mathrm{CXT} \rangle$ that they vary. Some methods vary the case specific parameters of a problem (OBS and CXT), others adjust the case independent parameters (SD). In this section we will discuss example techniques from each of these categories. If we want to use these techniques for computing approximate diagnoses, we also need to know their relative strengths: does one approximation also compute all solutions of another, are all solutions of one approximation larger than those of another, etc. We shall also discuss examples of such relations that hold between the approximation techniques that we present.

5.1 Adjusting the diagnostic relation

This class of approximations is based on adjusting various aspects of the relation we demand between observed values, the causal network and possible explanations. An example of such a relation is (3)–(4) above, and we will use this as our example.

Adjusting observation categories

Variations of (3)–(4) can be obtained by changing the construction of OBS^+ and OBS^- from OBS. The example (3)–(4) from Sec. 3 stated:

(OBS1): $\mathrm{OBS}^+ = \mathrm{OBS}$ and $\mathrm{OBS}^- = \{\neg O_i | O_i \in \mathcal{O} \backslash \mathrm{OBS}\}$.

We can obtain over-diagnoses by dropping the consistency demand (4):

(OBS2): $\mathrm{OBS}^+ = \mathrm{OBS}$ and $\mathrm{OBS}^- = \emptyset$,

Alternatively, we could drop (3). Under the following definition of OBS^- we obtain something very close to consistency-based diagnosis [Rei87].

(OBS3): $\mathrm{OBS}^+ = \emptyset$ and $\mathrm{OBS}^- = \{\neg O_i | O_i \in \mathrm{OBS}\}$

[CT92] define a whole variety of diagnostic relations in this way, and prove theorems about their inclusion relations. Among other things, they propose to partition OBS into normal and abnormal observations: $\mathrm{OBS} = \mathrm{OBS}_A \cup \mathrm{OBS}_N$, and to take

(OBS4): $\mathrm{OBS}^+ = \mathrm{OBS}_A$ and $\mathrm{OBS}^- = \mathrm{OBS}_N$.

This has the effect that a solution explains all the abnormal observations and does not contradict any of the normal observations.

A final variation can be obtained as follows. (OBS1) Assumes that any value outside OBS^+ must be taken as negative. This very strong closed world assumption can be relaxed by allowing OBS to contain negative as well as positive literals, and then define:

(OBS5): $\text{OBS}^+ = \{O_i \in \text{OBS}\}$ and $\text{OBS}^- = \{\neg O_i \in \text{OBS}\}$.

When computing approximate diagnoses, it is important to know the relative strengths of the different diagnostic relations defined in this way. For the cases described above, we can prove:

Property:

$$\begin{array}{ccccc} & & & & \text{OBS3} \\ & & & \nearrow & \\ \text{OBS1} & \rightarrow & \text{OBS2} & \rightarrow & \text{OBS4} \\ & & & \searrow & \\ & & & & \text{OBS5} \end{array}$$

Adjusting the provability relation

Almost all definitions of diagnostic relations are expressed in terms of the standard provability relation $\vdash$. We can therefore adjust DR by using a non-standard definition of $\vdash$ in (3)–(4).

For instance, we may want to approximate $\vdash$ by a relation that is easier to compute. An example of this occurs when we allow negative literals in OBS and then use negation as failure (written $\vdash_{NAF}$). Under the assumption that CN is Horn (as everywhere in this paper), we can then show the following:

Property:

If DR and DR' are defined as in (3)–(4) but using $\vdash_{NAF}$ instead of $\vdash$, and they satisfy $\text{OBS} = \text{OBS}^+ \cup \text{OBS}^-$

then DR(DP,SOL) *iff* DR'(DP,SOL) for any DP and SOL.

This states that when approximating $\vdash$ by $\vdash_{NAF}$, many variations OBSi from the previous section are no longer useful. For the examples from the previous section in particular we would have OBS2 $\leftrightarrow$ OBS4 $\leftrightarrow$ OBS5.

An altogether different approximation is inspired by [CS91]. They define approximate deduction relations which are sound (but incomplete) or complete (but unsound) approximations of $\vdash$ and $\not\vdash$. These approximations can be substituted in (3) and (4). In this way, we can reduce both the number and the size of the solutions obtained.

Adjusting the selection criterion

Almost every notion of diagnosis described in the literature contains a criterion to select the most prefered solution from among all logically possible ones. This is often formulated as selecting the minimal elements of some ordering on the solutions. By strengthening this selection criterion, we obtain less solutions.

The most common orderings in the literature are syntactic properties such as set-inclusion (we write $\subset$-min) or cardinality ordering (#-min). Both #-min and $\subset$-min are versions of Occam's Razor, but in general #-min only makes sense if the elements of the minimised set can be assumed to be independent.

Besides these syntactic orderings, we can also select solutions on more semantic grounds. A simple version of this is to order the elements of $\mathcal{S}$ or $\mathcal{L}$ (e.g. based on their urgency) and then to extend this to an ordering on sets of these elements such as STATES or LBLS.

In [vHtT94], we have proposed a selection mechanism (which we shall write P-min) based on annotating the causal links in CN with preference conditions, and then to select those solutions that make maximal use of preferred links.

In order to investigate more sophisticated selection mechanisms, we need a calculus of minimalities. In [tTvH94] we have proposed the following operations. If $\leq_1$ and $\leq_2$ are two orderings, then we can construct the intersection and union of their minimal elements (written $(\leq_1 \,\&\, \leq_2)$ and $(\leq_1 \mid \leq_2)$), or we can construct the lexicographic combination of $\leq_1$ and $\leq_2$ (written $(\leq_1; \leq_2)$). This is useful since minimal elements under $(\leq_1; \leq_2)$-min are guaranteed to exists whenever they exist under $\leq_1$ and $\leq_2$, which does not hold for $(\leq_1 \,\&\, \leq_2)$-min.

Using this notation, we can investigate how combinations of orderings behave. For instance, for #-min, $\subset$-min and P-min we have:

Property:

$$\begin{array}{ccccccc} \#\&\text{P-min} & \rightarrow & \#;\text{P-min} & \rightarrow & \underline{\#\text{-min}} & & \\ & \searrow & & \searrow & & \searrow & \\ & & \subset\&\text{P-min} & \rightarrow & \underline{\subset;\text{P-min}} & \rightarrow & \underline{\subset\text{-min}} \end{array}$$

We can also show under which circumstances combinations of minimalities have an approximating effect. The following shows that this is not always the case:

Property:

$$\begin{array}{ll} \text{if} & \leq_1\text{-min} \rightarrow \leq_2\text{-min} \\ \text{then} & \leq_1\text{-min} = (\leq_1 \,\&\, \leq_2)\text{-min} = (\leq_1; \leq_2)\text{-min} \\ \text{and} & \leq_2\text{-min} = (\leq_1 \mid \leq_2)\text{-min}. \end{array}$$

5.2 Adjusting the diagnostic problem

Adjusting the case-specific part

When the exact solutions for a given problem do not suffice, we may be better of with the solutions for a slightly adjusted version of the problem. We now discuss techniques that yield solutions for problems whose case-specific part (CXT,OBS) are approximations of those in the original problem.

The most obvious possibility is to manipulate the set OBS directly. We can impose a semantic ordering on $\mathcal{O}$ (e.g. based on the danger of the observations). If solutions SOL to $\text{DR}(\langle \text{SD}, \text{OBS}, \text{CXT} \rangle, \text{SOL})$ do not suffice, than we can create under-diagnoses by removing elements from OBS which are low in this ordering.

We can also try to find syntactic reasons for removing elements from OBS, such as ignoring aspecific symptoms (symptoms which can be caused by many different causes).

Approximate problems can also be found by manipulating the STATES involved in the exact solution of the original problem. An example of this is a method to ensure a complete diagnosis (see Sec. 4), by simply taking the union of all competing solutions. By a similar method we can obtain a sound approximation by taking the intersection over all competing solutions.

Many measures for making a diagnosis coherent also fall in this category. If a set is not coherent because only a few states are missing or superfluous, then we can add or remove such states from the solution, particularly if the offending states are low in some ordering of importance on $\mathcal{S}$. Besides such a semantic ordering, we may use syntactic measures on the importance of a state for the coherence of STATES, for instance by counting in how many labels a given state plays a role. Besides orderings on $\mathcal{S}$, orderings on $\mathcal{L}$ can also be used to decide how to make a diagnosis coherent.

Adjusting the case-independent part

The final category of techniques for computing approximations concerns permanent adjustments to the case-independent knowledge encoded in the causal network and the diagnostic labels. Due to space limitations, we will not discuss any details here, but only mention that an example would be a technique for specialising the network for certain classes of observations (e.g. common complaints, or dangerous ones), by applying techniques like partial evaluation, known from program transformation.

6 Related work

[PP91] motivates that the diagnostic problem, in most formal theories of diagnosis is incomplete, because the observation and hyptoheses phases are only considered and the treatment phases is not considered at all. As result these diagnosis approaches ignore utility considerations totally. This corresponds what we explain as the choice of an appropriate definition of diagnosis depends on the purpose for which diagnosis is performed.

Struß[Str92] was perhaps the first to consider the problem of choosing an appropriate notion of diagnosis as part of solving a diagnostic problem. His work lead Bötcher and Dressler to design the system Magellan [BD93], which explicitly models assumptions about the notion of diagnosis it uses, and can choose a different notion by changing these assumptions. Junker [Jun91] also shows a way of representing the assumptions underlying various diagnostic notions, and defines partial orders between them. His work can be seen as an alternative formulation of diagnostic assumption. All of this work deals mainly with consistency based diagnosis, whereas our notions of diagnosis also easily accomodate for abductive diagnosis.

We discussed a number of techniques for obtaining approximations, but our examples are by no means exhaustive. Two papers that present other techniques are [PEB94] which discusses various forms of abductive reasoning (instances of DR in our terminology), and [Bos94] which discusses approximations obtained by making abstractions of the system description SD.

7 Conclusion and future work

In this paper we have argued that diagnosis should not be seen as solving a problem with a unique definition, but rather that there exists a whole space of reasonable notions of diagnosis, and that these notions can be seen as mutual approximations (Sec. 2). We have presented a number of reasons for choosing among different notions of diagnosis (Sec. 4). We also presented an exhaustive categorisation of techniques that can be employed to obtain approximations (Sec. 5), as well as a number of specific example techniques for each category. We also showed that it is possible to characterise the relations between the approximations obtained by these techniques.

The perspective on diagnosis taken in this paper opens up a number of new questions. Of course, many additional techniques can be defined, and some of our definitions are only very simple first versions. However, for our techniques to become really useful, more fundamental questions also need to be answered. We must investigate the applicability criteria which determine when each approximation technique should be used. Furthermore, the techniques together with their applicability criteria should be incorporated in a computational architecture that can exploit these notions when performing diagnosis. Both these topics are the subject of our current research.

References

[BD93] Claudia Böttcher and Oskar Dressler. Diagnosis process dynamics: Holding the diagnostic trackhound in leash. In *Proceedings IJCAI–93*, 1993.

[Bos94] A. Bos. Can abstractions be used to speed up diagnostic reasoning ? In *Fifth International Workshop on Principles of Diagnosis (DX-94)*, 1994.

[CS91] M. Cadoli and M. Schaerf. Approximate entailment. In *Proceedings of 2nd Italian AI Conference AI*IA'91*, pages 68–77. Springer Verlag, Lecture Notes in AI, No. 549, 1991.

[CT90] L. Console and P. Torasso. Hypothetical reasoning in causal models. *Int. J. of Intelligent Systems*, 5(1):83–124, 1990.

[CT92] L. Console and P. Torasso. A spectrum of logical definitions of model-based diagnosis. In L. Console, J.H. de Kleer, and W.C. Hamscher, editors, *Readings in Model-based Diagnosis*. Morgan Kaufmann, 1992.

[Jun91] U. Junker. Generating diagnoses by prioritized defaults. In *Second International Workshop on Principles of Diagnosis*, pages 101–110, Milano, October 1991.

[PEB94] C. Preist, K. Eshghi, and B. Bertolino. Consistency-based and abductive diagnoses as generalised stable models. *Annual of Math. and AI special issues on Model-based diagnosis*, 11, 1994.

[PNdK94] W.M. Post, M. Neerincx, P. deGreef, and R.W. Koster. A design method for cognitive support applied to emergency call handling. Submitted, 1994.

[PP91] G.M. Provan and D. Poole. The utility of consistency–based diagnostic techniques. In *KR91 Principles of knowledge representation and reasoning*, Cambridge 1991.

[Rei87] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–96, 1987.

[Str92] P. Struss. Diagnosis as a process. In L. Console, J.H. de Kleer, and W.C. Hamscher, editors, *Readings in Model-based Diagnosis*. Morgan Kaufmann, 1992.

[tTvH94] A. ten Teije and F. van Harmelen. An extended spectrum of logical definitions for diagnostic systems. In *Proceedings of DX-94 Fifth International Workshop on Principles of Diagnosis*, 1994.

[vHtT94] F. van Harmelen and A. ten Teije. Using domain knowledge to select solutions in abductive diagnosis. In *ECAI'94*, pages 652–656, 1994.