Modelling Interactions for Diagnosis

Peter LUCAS
Department of Computer Science, Utrecht University
P.O. Box 80.089, 3508 TB Utrecht
The Netherlands
E-mail: lucas@cs.ruu.nl

Abstract

Model-based diagnosis concerns the development of diagnostic knowledge-based systems based on detailed domain models. Typically, such domain models include knowledge of causal, structural, and functional interactions among modelled objects. Various formal theories have been proposed in the literature to capture model-based diagnosis. In this paper, a new set-theoretical framework for the analysis of model-based diagnosis is presentented. Basically, the framework distinguishes between an interpretation of a specification of knowledge for the purpose of diagnosis, called an 'evidence function,' and an interpretation of this evidence function with respect to hypotheses and sets of observed findings, yielding diagnoses. This second interpretation is carried out by partial functions, called 'notions of diagnosis.' This set-theoretical framework offers a simple, yet powerful tool for comparing existing notions of diagnosis, as well as for proposing new notions of diagnosis.

Keywords: model-based diagnosis, formal methods.

1. INTRODUCTION

In recent years, model-based diagnosis has become a popular approach to building diagnostic systems in both technical (cf. [1], [5]), and nontechnical fields, such as medicine (cf. [6]). The model-based approach advocates the construction of knowledge-based systems based on models of a problem domain. Usually, such models describe the structural and functional interactions between components of a physical system, or the causal interactions between elements in a domain.

Accompanying research into the formal aspects of diagnosis has yielded much insight into the nature of the diagnostic process. Generally, two directions of research can be distinguished: (1) *consistency-based diagnosis*, which basically provides a theory of diagnosis for models of normal structure and behaviour [7], [10], and (2) *abductive diagnosis*, which focusses on causal models of abnormal behaviour [3]. As has been shown, consistency-based diagnosis can be extended to deal with fault models [7], and abductive diagnosis can be extended to models of normal behaviour [4]. Thus, these theories of diagnosis may both be used to lay out a spectrum of definitions of diagnosis.

However, the conclusion that there is not a unique way to characterize diagnosis raises questions concerning the assumptions underlying consistency-based and abductive diagnosis. Does the logical notion of consistency provide an appropriate basis for formalizing various notions of diagnosis, and similarly, is logical implication the proper way to formalize cause-effect and other relationships between defects and observable findings in abductive diagnosis?

In this paper, it is argued that the formalization of diagnosis requires the modelling of interactions at two levels of specification: interactions between defects, expressed by a mapping from defects to observable findings, and the interpretation of observed findings in the context of such a mapping. A set-theoretical semantic framework to express these aspects of diagnosis is proposed. Medicine and simple logic circuits are taken as example domains.

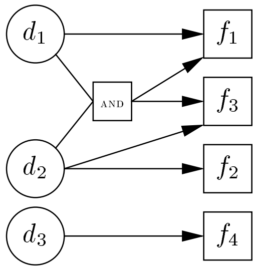

Figure 1. Nonmonotonic interaction between disorders.

2. THE REPRESENTATION OF INTERACTIONS

A. Causal interactions

Consider a medical diagnostic problem, where a patient may have Cushing's disease – a disease caused by a brain tumour producing hyperfunctioning of the adrenal glands –, pulmonary infection and iron-deficiency anaemia. We shall not enumerate all signs and symptoms causally associated with these medical problems; it suffices to note that moon face is a sign associated with Cushing's disease, fever and dyspnoea (shortness of breath) are associated with pulmonary infection, and low levels of serum iron are characteristic for iron-deficiency anaemia. However, in a patient in whom Cushing's disease and pulmonary infection coexist there usually is no fever. This indicates that there exists an interaction between the two disorders, Cushing's disease and pulmonary infection, that is nonmonotonic, i.e. the co-occurrence of the two disorders produces fewer findings than the union of their associated observable findings. Figure 1 depicts this simple problem, where it is assumed that:

d_1 = Cushing's disease
d_2 = pulmonary infection
d_3 = iron-deficiency anaemia
f_1 = moon face
f_2 = fever
f_3 = dyspnoea
f_4 = low serum iron

Interactions among disorders can be expressed by means of a mapping of sets of disorders to sets of observable findings. Such a mapping will be called an evidence function. Since the term 'disorder' is not used in technical domains, where instead the term 'fault' is commonly employed to indicate device problems, the term '*defect*' will be used in the following to denote both disorders in medicine and faults in technical devices.

More formally, let $\Sigma = (\Delta, \Phi, e)$ be a *diagnostic specification*, where Δ denotes a set of defects, and Φ denotes a set of findings. Positive defects d (findings f) and negative defects $\neg d$ (findings $\neg f$) denote *present* defects (findings) and *absent* defects (findings), respectively. It is assumed that $\neg \circ \neg = \iota$, where ι is the identity function. If a defect d or a finding f is not included in a set, it is assumed to be unknown. Let a set X_P denote a set of positive elements, and let X_N denote a set of negative elements, such that X_P and X_N are disjoint. It is assumed that $\Delta = \Delta_P \cup \Delta_N$ and $\Phi = \Phi_P \cup \Phi_N$. The power set of a set S is denoted by $\wp(S)$. Now, an *evidence function* e is a mapping

$$e : \wp(\Delta) \to \wp(\Phi) \cup \{\bot\}$$

such that:

(1) for each $f \in \Phi$ there exists a set $D \subseteq \Delta$ with $f \in e(D)$ or $\neg f \in e(D)$ (and possibly both);
(2) if $d, \neg d \in D$ then $e(D) = \bot$;
(3) if $e(D) \neq \bot$ and $D' \subseteq D$ then $e(D') \neq \bot$.

If $e(D) \neq \bot$, it is said that $e(D)$ is the set of *observable findings* for D; otherwise, it is said that D is inconsistent.

According to the definition above, we may have that both $f \in e(D)$ and $\neg f \in e(D)$, which simply means that these findings may alternatively occur given the combined occurrence of the defects in the set D. In some domains it might hold that if $e(\{d\}) = e(\{d'\})$, it follows that $d = d'$, i.e. the defects d and d' are taken as synonyms for the same defect. An evidence function is not assumed to be injective in general, because for non-singleton sets $D, D' \subseteq \Delta$, $D \neq D'$, it is not precluded that $e(D) = e(D')$. It is also not precluded that sets of defects may have several findings in common; thus, the sets $e(D)$ and $e(D')$, $D \neq D'$, need not be disjoint.

For the medical knowledge depicted in Figure 1, it holds, among others, that:

$$\begin{aligned} e(\{d_1\}) &= \{f_1\} \\ e(\{d_2\}) &= \{f_2, f_3\} \\ e(\{d_1, d_2\}) &= \{f_1, f_3\} \end{aligned}$$

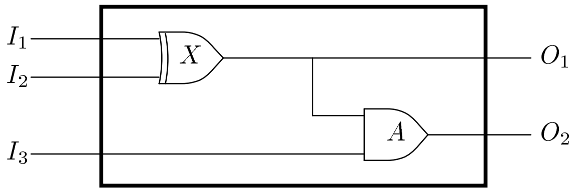

Figure 2. Logic circuit.

The property $e(\{d_1, d_2\}) \not\supseteq e(\{d_1\}) \cup e(\{d_2\})$ formally expresses that the interaction between d_1 and d_2 is nonmonotonic.

B. Functional behaviour

As discussed in Section 1, there are, in addition to diagnostic systems that incorporate causal knowledge, other types of diagnostic systems containing knowledge of structure and behaviour. This type of knowledge is usually employed for diagnosing device problems, where the behaviour of the device is observed by means of input and output signals. Consider the logic circuit depicted in Figure 2. The circuit consists of an XOR (exclusive OR) gate X and an AND gate A. The presence of a defect in X is denoted by x; the absence of a defect in X is denoted by $\neg x$. A similar notation is employed to denote the presence or absence of a defect concerning gate A. The three inputs signals to the circuit are indicated by I_1, I_2 and I_3; O_1 and O_2 denote the two output signals. If $I_j = 1$, this will be denoted by i_j; an input equal to $I_j = 0$ will be denoted by $\neg i_j$. A similar convention is adopted for the output signals O_k. It is supposed that a defective gate produces an output signal that is complementary to the correct output signal. Suppose that the input signals to the circuit are $i_1, \neg i_2$ and i_3. Now, the output signals are represented as observable findings, and a component for which the presence or absence of a defect is unknown, is taken into account by assuming that the component is either defective or nondefective. Note that this description concerns both the structure as well as the normal and abnormal behaviour of the device. The following evidence function (only values for consistent sets of defects are provided) corresponds to the description above:

$$\begin{aligned} e'(\{x, a\}) &= \{\neg o_1, o_2\} \\ e'(\{\neg x, a\}) &= \{o_1, \neg o_2\} \\ e'(\{x, \neg a\}) &= \{\neg o_1, \neg o_2\} \\ e'(\{\neg x, \neg a\}) &= \{o_1, o_2\} \\ e'(\{x\}) &= \{\neg o_1, o_2, \neg o_2\} \\ e'(\{\neg x\}) &= \{o_1, o_2, \neg o_2\} \\ e'(\{a\}) &= \{o_1, \neg o_1, o_2, \neg o_2\} \\ &= e'(\{\neg a\}) \end{aligned}$$

Meaning	Definition in terms of observable findings
causality: the defects D cause D'	$e(D') \subseteq e(D)$ and $e(D) = e(D \cup D')$
correlation: defects d and d' are correlated	$e(\{d\}) = e(\{d'\}) = e(\{d, d'\})$ and $e(\{\neg d\}) = e(\{\neg d'\}) = e(\{\neg d, \neg d'\})$
category: d' is a category for defects in D	$e(\{d'\}) = \bigcup_{d \in D} e(\{d\})$
augmentation	$e(D) \supset \bigcup_{D' \subset D} e(D')$
cancellation (fault masking)	$e(D) \subset \bigcup_{D' \subset D} e(D')$
exclusion	$e(D) = \perp$
complementation	if $e(\{d_1, \ldots, d_n\}) = \{f_1, \ldots, f_m\}$ then $e(\{\neg d_1, \ldots, \neg d_n\}) = \{\neg f_1, \ldots, \neg f_m\}$

Table 1
LOCAL INTERACTIONS AMONG DEFECTS.

$$= \quad e'(\varnothing)$$

For example, $e'(\{x\}) = \{\neg o_1, o_2, \neg o_2\}$ indicates that when the XOR gate is defective, and it is unknown whether or not the AND gate is defective, then the first output signal $O_1 = 0$ and the second output signal O_2 may be either 0 or 1, depending on whether the AND gate is defective or not.

Knowing which findings are observable for present or absent defects is essentially the qualitative information that is required for diagnosis.

3. SOME PROPERTIES OF EVIDENCE FUNCTIONS

Various semantic properties of a domain for which a diagnostic system must be built can be expressed precisely in terms of evidence functions. In Table 1 some possible *local* interactions among defects, i.e. interactions that only hold for some $D \subseteq \Delta$, are enumerated. Based on the particular semantics underlying a knowledge base (left column), an interpretation that respects that meaning in terms of observable findings is defined (right column). Examples of local interactions are:

- *Causality*: the diagnostic view of knowledge of the sort 'the set of defects D *causes* the set of defects D'' as used in abductive diagnosis (cf. [3]).
- *Correlation*: if the defects d and d' are correlated, then if d has occurred, then d' occurs as well, whereas, if d is absent, d' is also absent, and vice versa.
- *Category*: a category gathers all findings of the defects with regard to which it is more general.
- *Augmentation*: the combined occurrence of two or more defects in the set D gives rise to new observable findings in addition to those associated with the individual elements, or proper subsets of D.
- *Cancellation* (also referred to as *fault masking*): the combined occurrence of two or more defects in the set D yields fewer observable finding when compared to the findings associated with the individual elements, or proper subsets of D.
- *Exclusion*: the combination of defects D cannot occur.
- *Complementation*: the observable findings associated with the absent defects $\neg d_1, \ldots, \neg d_n$ are the complements of those associated with the presence of those.

We may also have that defects exhibit no interactions at all, which is a *global* property, expressed as follows:

$$e(D) = \bigcup_{d \in D} e(\{d\})$$

for each consistent set $D \subseteq \Delta$. Monotonicity in terms of set-inclusion is also a global property of the evidence function, which holds for many forms of causality and normal behaviour of devices expressed in terms of evidence functions. For example, for the logic circuit in Figure 2 it holds that

$$\forall D, D' \subseteq \Delta : D \subseteq D' \Rightarrow e'(D) \supseteq e'(D')$$

i.e. e' is *monotonically decreasing*.

An evidence function requires an exponential number of interactions to be specified. However, providing a partial specification of interactions may suffice when it can be assumed that the remaining values of e can be computed according to some computation rule. For example, assuming that all values of the medical evidence function e above can be obtained by taking the union of consistently specified function values for sets of defects that are maximal subsets of a given set of defects, yields the following partial specification $\tilde{e}$ of the evidence function e (cf. Figure 1):

$$\tilde{e}(D) = \begin{cases} \{f_1\} & \text{if } D = \{d_1\} \\ \{f_2, f_3\} & \text{if } D = \{d_2\} \\ \{f_4\} & \text{if } D = \{d_3\} \\ \{f_1, f_3\} & \text{if } D = \{d_1, d_2\} \\ \varnothing & \text{if } D = \{d\}, d \in \Delta_N \end{cases}$$

For example, it holds that $e(\{d_1, d_3\}) = \tilde{e}(\{d_1\}) \cup \tilde{e}(\{d_3\})$. Note that the generated evidence function displays cancellation of observable findings (cf. Table 1), because

$$e(\{d_1, d_2\}) \subset e(\{d_1\}) \cup e(\{d_2\})$$

Using another computation rule, it is also possible to provide a partial specification of the evidence function e' for the logical circuit.

4. NOTIONS OF DIAGNOSIS

As has been shown, an evidence function can be viewed as a semantic interpretation of a knowledge base, containing for example causal or functional knowledge, in terms of expected evidence for the combined occurrence of defects. To employ an evidence function for the purpose of diagnosis, it must be interpreted with respect to the actually observed findings. The interpretation of an evidence function and the observed findings that is adopted, can be viewed as a notion of diagnosis applied to solve the diagnostic problem at hand.

More formally, let $\mathcal{P} = (\Sigma, E)$ be a *diagnostic problem*, where $E \subseteq \Phi$ is the set of *observed findings*; it is assumed that if $f \in E$ then $\neg f \notin E$, i.e. contradictory observed findings are not allowed. Let R_Σ denote a notion of diagnosis R applied to Σ, then a mapping

$$R_{\Sigma, e_{|H}} : \wp(\Phi) \to \wp(\Delta) \cup \{u\}$$

will either provide a diagnostic solution for a diagnostic problem $\mathcal{P}$, or indicate that no solution exists, denoted by u (undefined). Here, H denotes a *hypothesis*, which is taken to be a set of defects ($H \subseteq \Delta$), and $e_{|H}$, called the *restricted evidence function* of e, is a restriction of e with respect to the power set $\wp(H)$:

$$e_{|H} : \wp(H) \to \wp(\Phi) \cup \{\bot\}$$

where for each $D \subseteq H$: $e_{|H}(D) = e(D)$. A restricted evidence function $e_{|H}$ can be thought of as the relevant part of a knowledge base with respect to a hypothesis H. An *R-diagnostic solution*, or *R-diagnosis* for short, with respect to a hypothesis $H \subseteq \Delta$, is now defined as the set

$$R_{\Sigma, e_{|H}}(E), \text{ where } R_{\Sigma, e_{|H}}(E) \subseteq H \text{ if a solution exists.}$$

In Figure 3, the idea underlying the definition of a notion of diagnosis R and diagnostic solution to a diagnostic problem is illustrated schematically.

A notion of diagnosis R provides the possibility to express interactions among defects and observed findings at the level of diagnosis, which we call dependencies. We may also have that a hypothesis can be split up into two subhypotheses, that can be examined independently:

$$R_{\Sigma, e_{|H \cup H'}}(E) = R_{\Sigma, e_{|H}}(E) \cup R_{\Sigma, e_{|H'}}(E)$$

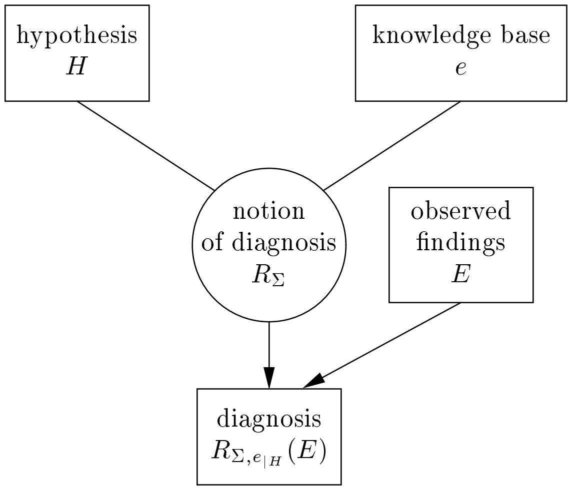

Figure 3. Schema of notion of diagnosis, diagnostic problem and solution.

with $R_{\Sigma, e_{|H \cup H'}}(E) \neq u$. This means that the diagnostic solution with respect to the hypothesis $H \cup H'$ is obtained as the union of the solutions for the two separately examined hypotheses H and H'. This is called the *independence assumption*. For many notions of diagnosis described in the literature, in particular for abductive diagnosis and consistency-based diagnosis, the independence assumption fails to hold.

To demonstrate how the definitions above can be employed, we consider a notion of diagnosis U, such that $U_{\Sigma, e_{|H}}(E) = H'$ if it holds that H' is the only subset of H such that $e_{|H}(H') \subseteq E$; otherwise, $H' = u$. This notion of diagnosis expresses that a diagnosis consists of a set of defects which, on the one hand, can account for at least part of all observed findings, and, on the other hand, every finding associated with the set of defects that is taken as a diagnosis has been observed. Furthermore, there is only one such subset of the given hypothesis H. Now, reconsider the medical example from Figure 1 with $H = \{d_1, d_2\}$. Some interesting diagnostic conclusions are: $U_{\Sigma, e_{|H}}(\{f_2, f_3\}) = \{d_2\}$, i.e. a patient with only fever and dyspnoea has pulmonary infection, $U_{\Sigma, e_{|H}}(\{f_1, f_2\}) = u$, i.e. there exists no diagnosis accounting for both moon face and fever as signs, and finally, $U_{\Sigma, e_{|H}}(\{f_1, f_3\}) = H$. In the first case, it is said that the hypotheses has been *adjusted*, in the second case, that the hypothesis H is *rejected*, and in the last case, that the hypothesis H has been *accepted*. This example demonstrates the flexibility of the approach.

5. ANALYSIS OF NOTIONS OF DIAGNOSIS FROM THE LITERATURE

Because the diagnostic formalism introduced above is meant to act as a framework, various notions of diagnosis known from the literature should be expressible in it. In this section, the expressive power of the framework is examined with respect to abductive and consistency-based diagnosis.

A. Abductive diagnosis

The formalization of diagnosis using causal domain models has been thoroughly studied by L. Console and P. Torasso [3], [4]. In their theory, the abnormal behaviour of a system is specified in terms of abnormal states (called defects in this paper) and resulting abnormal findings; normal findings, however, may also be included. *Strongly* causal relationships between defects, and between defects and observable findings, are expressed by logical implications of the form $d_1 \wedge \cdots \wedge d_n \rightarrow d'$ and $d_1 \wedge \cdots \wedge d_n \rightarrow f$, respectively. Consider, for example, the *causal specification* $\mathcal{C} = (\Delta, \Phi, \mathcal{R})$, with $\Delta_P = \{d_1, d_2, d_3\}$, $\Phi_P = \{f_1, f_2, f_3\}$, where specific causal knowledge with respect to defects and observable findings is expressed by the following set of *abnormality axioms* $\mathcal{R}$:

$$
\begin{aligned}
&d_1 \rightarrow f_1\\
&d_1 \rightarrow f_2\\
&d_2 \rightarrow d_1\\
&d_3 \rightarrow f_3
\end{aligned}
$$

Now, let $\mathcal{A} = (\mathcal{C}, E)$ be an abductive diagnostic problem, with $E = \{f_1, f_2\}$ a set of observed findings. Then, a set of defects $H \subseteq \Delta$ is called a *diagnosis* of $\mathcal{A}$ iff:

(1) $\forall f \in E : \mathcal{R} \cup H \vDash f$ (*covering condition*);
(2) $\forall f \in E^c : \mathcal{R} \cup H \nvDash \neg f$ (*consistency condition*)

where E^c, the set of observable findings assumed to be absent, is defined in terms of E as follows:

$$E^c = \{\neg f \in \Phi \mid f \in \Phi_P, f \notin E\}$$

Here, we have that $E^c = \{\neg f_3\}$, and, thus, $H = \{d_1, d_2\}$ is a diagnosis for $\mathcal{A}$, because the covering and consistency conditions are satisfied.

This form of abductive diagnosis can be translated into our framework in a straightforward way. For the axioms $\mathcal{R}$ above, a partial specification $\tilde{e''}$ of the resulting evidence function e'', where again unspecified function values are obtained by taking the union of specified ones, is:

$$
\tilde{e''}(D) = \begin{cases}
\{f_1, f_2\} & \text{if } D = \{d_1\}\\
\{f_1, f_2\} & \text{if } D = \{d_2\}\\
\{f_3\} & \text{if } D = \{d_3\}\\
\bot & \text{if } D = \{\neg d_1, d_2\}\\
\varnothing & \text{if } D = \{\neg d_i\}, i = 1, 2
\end{cases}
$$

yielding a diagnostic specification $\Sigma = (\Delta, \Phi, e)$.

Abductive diagnosis as defined above can now be defined as a notion of diagnosis. The corresponding notion of diagnosis is called the notion of *strong-causality diagnosis* (SC). It is defined as follows:

$$
\mathrm{SC}_{\Sigma, e_{|H}}(E) = \begin{cases}
H & \text{if } e_{|H}(H) = E\\
u & \text{otherwise}
\end{cases}
$$

i.e. it is necessary that all observable findings $e(H)$ are observed to accept an hypothesis H.

For the diagnostic problem $\mathcal{P} = (\Sigma, E)$, with $E = \{f_1, f_2\}$, we find:

$$\mathrm{SC}_{\Sigma, e_{|\{d_1, d_2\}}}(\{f_1, f_2\}) = \{d_1, d_2\}$$

Note that for $E' = \{f_1\}$ no abductive diagnosis exists. Indeed, it holds that $\mathrm{SC}_{\Sigma, e_{|H}}(E') = u$ for $E' = \{f_1\}$ and every consistent $H \subseteq \Delta$.

A notion of *weak* causality [3] is arrived at by the addition of assumption literals α to the individual abnormality axioms. This way, it can be expressed that a causal relation is uncertain. Reconsider the abductive problem $\mathcal{A} = (\mathcal{C}, E)$ above, where assumption literals are added to the individual axioms, yielding the causal specification $\mathcal{C}' = (\Delta', \Phi, \mathcal{R}')$, with $\mathcal{R}'$ equal to:

$$
\begin{aligned}
&d_1 \wedge \alpha_1 \rightarrow f_1\\
&d_1 \wedge \alpha_2 \rightarrow f_2\\
&d_2 \wedge \alpha_3 \rightarrow d_1\\
&d_3 \wedge \alpha_4 \rightarrow f_3
\end{aligned}
$$

The resulting evidence function is again e'' as defined above. The diagnostic interpretation of this evidence function, however, differs. To this end, a distinction is made between an abductive *solution* – a set of defects and assumption literals for which the covering and consistency conditions are satisfied –, and an (abductive) *diagnosis*, the set of all defects included in an abductive solution.

The notion of diagnosis that corresponds to abductive diagnosis, with weakly causal relations as introduced above, is called the notion of *weak-causality diagnosis*, denoted by WC. It is defined as follows:

$$
\mathrm{WC}_{\Sigma, e_{|H}}(E) = \begin{cases}
H & \text{if } e_{|H}(H) \supseteq E\\
u & \text{otherwise}
\end{cases}
$$

For example, the set $H = \{d_1, \alpha_1, \alpha_2\}$ is an abductive solution to $\mathcal{A}' = (\mathcal{C}', E)$, because the covering and consistency conditions are satisfied; the associated dagnosis is $D = \{d_1\}$. We also have that $\mathrm{WC}_{\Sigma, e_{|\{d_1\}}}(E) = \{d_1\}$.

Weak and strong causality diagnosis can also be combined to obtain a notion of diagnosis that combines these two different interpretations of causal knowledge.

B. Consistency-based diagnosis

In consistency-based diagnosis, as proposed in [10] and [7], knowledge concerning structure and behaviour of a device is represented as a triple $\mathcal{S} = (\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$, called a *system*, where

- SD denotes a finite set of formulae in first-order predicate logic, specifying normal structure and behaviour, called the *system description*;
- COMPS denotes a finite set of constants in first-order logic, denoting the *components* (elements)

of the system;

- OBS denotes a finite set of formulae in first-order predicate logic, denoting *observations*, i.e. observed findings.

It is, in principle, possible to specify normal as well as abnormal (faulty) behaviour within a system description SD.

A consistency-based diagnosis is defined as an assignment of either a positive literal Abnormal(c) or a negative literal $\neg$Abnormal(c) to each $c \in$ COMPS, i.e.

$$D = \{\text{Abnormal}(c) \mid c \in C\} \cup \{\neg\text{Abnormal}(c) \mid c \in \text{COMPS} \backslash C\}$$

where $C \subseteq$ COMPS, such that

$$\text{SD} \cup \text{OBS} \cup D$$

is satisfiable (the *consistency condition*).

Again, the notion of diagnosis can be defined in terms of our framework. The resulting notion of *consistency-based diagnosis*, denoted by CB, is defined as follows:

$$\text{CB}_{\Sigma, e_{|H}}(E) = \begin{cases} H & \text{if } \forall f \in E : f \in e_{|H}(H) \vee \\ & \quad \neg f \notin e_{|H}(H) \\ u & \text{otherwise} \end{cases}$$

For example, for the logic circuit in Figure 2 we have that $\text{CB}_{\Sigma, e_{|\{x,a\}}}(\{\neg o_1, o_2\}) = \{x, a\}$, which is analogous to the diagnosis

$$D = \{\text{Abnormal}(x), \text{Abnormal}(a)\}$$

obtained by the corresponding logical definition of consistency-based diagnosis.

We have investigated the expressive power of the framework for other notions of diagnosis in the literature as well (cf. [8]).

6. COMPARISON TO RELATED WORK

Above we have introduced a quite general framework to express *static* aspects of diagnosis, i.e. without taking diagnostic problem-solving strategies into account. The framework supports two different views. On the one hand, given some intuitively appealing interpretation of an evidence function, a notion of diagnosis can be designed (or selected) that adheres to that meaning as closely as possible. On the other hand, applying a particular notion of diagnosis to solve a diagnostic problem implies that a particular (diagnostic) meaning is given to the associated evidence function.

The framework, which is inspired by the work on abductive diagnosis by Reggia et al. ([9]) and Bylander et al. ([2]), differs in several respects from the diagnostic frameworks based on logic [4], [7], [10], [11]. Firstly, the logical notions of diagnosis proposed in the literature have been designed in close connection with specific domain models, such as causal models or models of structure and behaviour. In contrast, in our framework, there is no intimate connection between the theory and any of the existing conceptual models of diagnosis. In fact, the meaning of a knowledge base, described by means of an evidence function e, is completely separated from its diagnostic use. Of course, it is usually desirable to define notions of diagnosis that closely mirror the meaning of a knowledge base. Secondly, where in the other frameworks, the modelled functional behaviour is usually monotonic (except when nonmonotonic logic is employed), due to the monotonicity of the employed logical entailment relation, monotonicity is no prerequisit in our framework.

In our framework, the properties of an evidence function follow from domain characteristics. Standard properties of evidence functions, as illustrated in Table 1, and notions of diagnosis express standard interpretations of domain knowledge. A limitation is that, as a tool for the semantical analysis of diagnosis, our framework is rather extensional in nature. This is in contrast to the more intensional nature of logic-based techniques for the analysis of diagnosis, such as used in defining consistency-based and abductive diagnosis, which allow for the separate specification of knowledge of structure and function.

REFERENCES

[1] A. Beschta, O. Dressler, H. Freitag, M. Montag, P. Struss (1993). DPNet – a second generation expert system for localizing faults in power transmission networks. In *Proceedings of the International Conference on Fault Diagnosis* (Tooldaig93), Toulouse, pp. 1019–1027.

[2] T. Bylander. D. Allemang, M.C. Tanner and J.R. Josephson (1992). The computational complexity of abduction. In *Knowledge Representation* (R.J. Brachman, H.J. Levesque and R. Reiter, eds.), pp. 25–60. Cambridge, Massachusetts: The MIT Press.

[3] L. Console, D. Theseider Dupré and P. Torasso (1989). A theory of diagnosis for incomplete causal models. *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pp. 1311–1317.

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

[5] P. Dague (1994). Model-based diagnosis of analog electronic circuits. *Annals of Mathematics and Artificial Intelligence*, **11**, 439–492.

[6] K.L. Downing (1993). Physiological applications of consistency-based diagnosis. *Artificial Intelligence in Medicine*, **5**, 9–30.

[7] J. de Kleer, A.K. Mackworth and R. Reiter (1992). Characterizing diagnoses and systems. *Artificial Intelligence*, **52**, 197–222.

[8] P.J.F. Lucas (1996). *Structures in Diagnosis: from theory to medical application.* PhD Thesis, Free University of Amsterdam.

[9] Y. Peng and J.A. Reggia (1990). *Abductive Inference Models for Diagnostic Problem Solving.* New York: Springer-Verlag.

[10] R. Reiter (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, **32**, 57–95.

[11] A. ten Teije and F. van Harmelen (1994). An extended spectrum of logical definitions for diagnostic systems. In *DX-94, 5th International Workshop on Principles of Diagnosis* (G.M. Provan, ed.), pp. 334–342.