Conflict-based Diagnosis: Adding Uncertainty to Model-based Diagnosis

Ildikó Flesch, Peter Lucas and Theo van der Weide
Department of Information and Knowledge Systems, ICIS
Radboud University Nijmegen, the Netherlands
Email: {ildiko,peterl,tvdw}@cs.ru.nl

Abstract

Consistency-based diagnosis concerns using a model of the structure and behaviour of a system in order to analyse whether or not the system is malfunctioning. A well-known limitation of consistency-based diagnosis is that it is unable to cope with uncertainty. Uncertainty reasoning is nowadays done using Bayesian networks. In this field, a conflict measure has been introduced to detect conflicts between a given probability distribution and associated data.

In this paper, we use a probabilistic theory to represent logical diagnostic systems and show that in this theory we are able to determine consistent and inconsistent states as traditionally done in consistency-based diagnosis. Furthermore, we analyse how the conflict measure in this theory offers a way to favour particular diagnoses above others. This enables us to add uncertainty reasoning to consistency-based diagnosis in a seamless fashion.

1 Introduction

Model-based diagnostic reasoning is concerned with the diagnosis of malfunctioning of systems, based on an explicit model of the structure and behaviour of these systems [Reiter, 1987]. In the last two decades, model-based diagnosis has become an important area of research with applications in various fields, such as software engineering [Köb and Wotawa, 2004] and the automotive industry [Struss and Price, 2003].

Basically, two types of model-based diagnosis are being distinguished in literature: (*i*) consistency-based diagnosis [Reiter, 1987], and (*ii*) abductive diagnosis [Console et al., 1990]. In this paper, we only deal with consistency-based diagnosis.

Consistency-based diagnosis generates diagnoses by comparing the predictions made by a model of structure and behaviour with the observations; it determines the behavioural assumptions under which predictions and observations are consistent. It is typically used for trouble shooting of devices that are based on a design [Genesereth, 1984].

A limitation of consistency-based diagnosis is that it is only capable of handling qualitative knowledge and unable to cope with the uncertainty that comes with many problem domains. This implies that an important feature of diagnostic problem solving is not captured in the theory. To solve this problem, de Kleer has proposed adding uncertainty to consistency-based diagnosis by specifying a joint probability distribution on all possible behavioural assumptions, taking these to be mutually independent [de Kleer, 1990]. One step further is a proposal by Kohlas et al. to adjust the probability distribution by excluding diagnoses that are inconsistent [Kohlas et al., 1998]. In both cases, consistency-based diagnosis and uncertainty reasoning are kept separate.

There have also been proposals to utilise Bayesian networks as a probabilistic framework for model-based diagnosis [Pearl, 1988]. Poole has proposed using consistency-based diagnosis to speed up reasoning in a Bayesian network [Poole, 1996]. Lucas has proposed a method to integrate consistency-based diagnosis into Bayesian network reasoning [Lucas, 2001].

None of the approaches above suggest determining and ordering diagnoses in a probabilistic manner, yet in a way similar to consistency-based diagnosis.

In this paper, we explore a probabilistic framework that models the structure and behaviour of logical diagnostic systems. The two major aims of our research were to develop a new probabilistic framework that

1. is capable of distinguishing between consistent and inconsistent states of a system and, therefore, allows determining diagnoses;

2. offers a way to favour particular diagnoses above others by means of a statistical measure.

The first aim is achieved by defining consistency and inconsistency probabilistically; the second aim is fulfilled by using the conflict measure from Bayesian network as such a statistical measure [Jensen, 2001].

The paper is organised as follows. In Section 2, the necessary basic concepts are defined. Subsequently, in Section 3, the definition of Bayesian diagnostic problems is given together with probabilistic definitions of consistency and inconsistency. Section 4 shows that the conflict measure is capable of ordering diagnoses. Finally, in Section 5 the results of this paper are summarised.

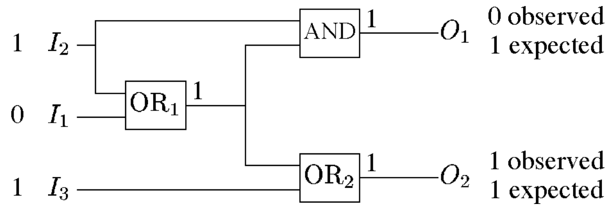

Figure 1: Example of a circuit.

2 Preliminaries

In this section, we provide a brief summary of the theories of consistency-based diagnosis and Bayesian networks.

2.1 Consistency-based Diagnosis

In the theory of consistency-based diagnosis [Reiter, 1987], the structure and behaviour of a system is represented by a *logical diagnostic system* $\mathcal{S}_L = (\mathrm{SD}, \mathrm{CMP})$, where

- SD denotes the *system description*, which is a finite set of logical formulae, specifying structure and behaviour;
- CMP is a finite set of constants, corresponding to the *components* of the system; these components can be faulty.

The system description consists of *behaviour descriptions*, and *connections*. A behavioural description is a formula specifying *normal* and *abnormal* (faulty) functionalities of the components. These normal and abnormal functionalities are indicated by *abnormality literals*. A connection is a formula of the form $i_c = o_{c'}$, where i_c and $o_{c'}$ denote the input and output of components c and c'.

A *logical diagnostic problem* is defined as a pair $\mathcal{P}_L = (\mathcal{S}_L, \mathrm{OBS})$, where $\mathcal{S}_L$ is a logical diagnostic system and OBS is a finite set of logical formulae, representing *observations*.

Adopting the definition from [de Kleer et al., 1992], a diagnosis in the theory of consistency-based diagnosis is defined as follows. Let Δ be the assignment of either a normal or an abnormal behavioural assumption to *each* component. Then, Δ is a *consistency-based diagnosis* of the logical diagnostic problem $\mathcal{P}_L$ iff the observations are consistent with both the system description and the diagnosis:

$$\mathrm{SD} \cup \Delta \cup \mathrm{OBS} \nvDash \bot.$$

Here, $\nvDash$ stands for the negation of the logical entailment relation, and $\bot$ represents a contradiction.

EXAMPLE 1 Consider Figure 1, which depicts an electronic circuit with one AND gate and two OR gates. Now, the output of the system differs from the one expected according to the simulation model, thus it gives rise to an inconsistency. One of the diagnoses, resolving the inconsistency, is to assume that the AND gate is functioning abnormally.

2.2 Bayesian Networks and Data Conflict

Let $P(X_V)$ denote a joint probability distribution of the set of *discrete* random variables X_V with finite set of indices V. Let $U, W, Z \subseteq V$ be mutually disjoint sets of indices. Then,

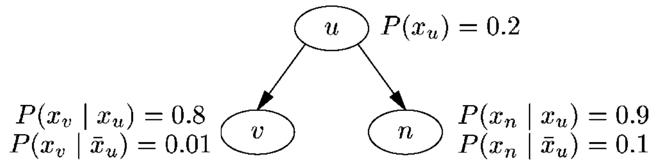

Figure 2: Example of a Bayesian network.

the set of random variables X_U is said to be *conditionally independent* of X_W *given* X_Z, if

$$P(X_U \mid X_W, X_Z) = P(X_U \mid X_Z). \tag{1}$$

A *Bayesian network* is a pair $\mathcal{B} = (G, P)$, where all independencies in the acyclic directed graph G are also contained in P, and P is factorised according to G as

$$P(X_V) = \prod_{v \in V} P(X_v \mid \pi_v), \tag{2}$$

where π_v denotes the random variables associated with the parent set of vertex v in the graph. In this paper, we assume that all random variables are binary; x_v stands for a positive value of X_v, whereas $\bar{x}_v$ denotes a negative value.

Bayesian networks specify particular probabilistic patterns that must be fulfilled by observations. Observations are random variables that obtain a value through an intervention, such as a diagnostic test. The set of observations is denoted by Ω. The *conflict measure* has been proposed as a tool for the detection of potential conflicts between observations and a given Bayesian network [Jensen, 2001], and is defined as:

$$\mathrm{conf}(\Omega) = \log \frac{P(\Omega_1)P(\Omega_2)\cdots P(\Omega_m)}{P(\Omega)}, \tag{3}$$

with $\Omega = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_m$.

The interpretation of the conflict measure is as follows. A zero or negative conflict measure means that the denominator is equally likely or more likely than the numerator. This is interpreted as that the joint occurrence of the observations is in accordance with the probabilistic patterns in P. A positive conflict measure, however, implies negative correlation between the observations and P indicating that the observations do not match P very well.

The interpretation of the conflict measure is illustrated by Example 2.

EXAMPLE 2 Consider the Bayesian network shown in Figure 2, which describes that stomach ulcer (u) may give rise to both vomiting (v) and nausea (n).

Now, suppose that a patient comes in with the symptoms of vomiting and nausea. The conflict measure then has the following value:

$$\mathrm{conf}(\{x_v, x_n\}) = \log \frac{P(x_v)P(x_n)}{P(x_v, x_n)} = \log \frac{0.168 \cdot 0.26}{0.1448} \approx -0.5.$$

As the conflict measure assumes a negative value, there is no conflict between the two observations. This is consistent with medical knowledge, as we do expect that a patient with stomach ulcer displays symptoms of both vomiting and nausea.

As a second example, suppose that a patient has only symptoms of vomiting. The conflict measure now obtains the following value:

$$\text{conf}(\{x_v, \bar{x}_n\}) = \log \frac{0.168 \cdot 0.74}{0.0232} \approx \log 5.36 \approx 0.7.$$

As the conflict measure is positive, there is a conflict between the two observations, which is again in accordance to medical expectations.

3 Probabilistic Diagnosis

The main aim of this section is to define a probabilistic theory that is related to consistency-based diagnosis.

In Section 3.1, the system description and the components of a logical diagnostic system are mapped to a probabilistic representation, defined along the lines of [Pearl, 1988] and [Poole, 1996]. This representation is called a Bayesian diagnostic system, which, together with the observations Ω, yield a Bayesian diagnostic problem. In Section 3.2, consistency and inconsistency are defined for Bayesian diagnostic problems.

3.1 Bayesian Diagnostic Problems

To start, we introduce some necessary notation. In the remaining part of this paper, the set CMP acts as the set of indices to the components of a diagnostic system. In this context, C denotes a subset of these components, whereas c indicates an individual component.

We are now in a position to define Bayesian diagnostic systems. In this formalism, the relations between the components are defined qualitatively by a graph and quantitatively by a probability distribution.

A *Bayesian diagnostic system*, defined as a pair $\mathcal{S}_B = (\mathcal{B}, \text{CMP})$, is obtained as the image of a logical diagnostic system $\mathcal{S}_L$, where: (*i*) $\mathcal{B} = (G, P)$ is a Bayesian network with acyclic directed graph G and joint probability distribution P of the set of random variables X_V; (*ii*) the acyclic directed graph $G = (V, E)$ is the image of the system description, with $V = O \cup I \cup A$; O are the *output vertices* corresponding to the outputs of components, I are *input vertices* corresponding to the inputs of components and A are *abnormality vertices* corresponding to abnormality literals. The set of arcs E results from the mapping of connections in SD.

According to the definition above, the set of random variables corresponds one-to-one to the set of vertices, thus $X_V \leftrightarrow V$. Figure 3 shows the graphical representation of the Bayesian diagnostic system corresponding to the circuit in Figure 1.

As we have already mentioned, we also need to add observations, thus inputs and outputs, to Bayesian diagnostic systems. It is generally not the case that the entire set of inputs and outputs of a system is observed. We, therefore, make the following distinction: observed system inputs and outputs will be denoted by I_S and O_S, whereas the remaining (non-observed) inputs and outputs are denoted by I_R and O_R. Clearly, $I = I_S \cup I_R$ and $O = O_S \cup O_R$.

Now, we are ready to define the notion of Bayesian diagnostic problem.

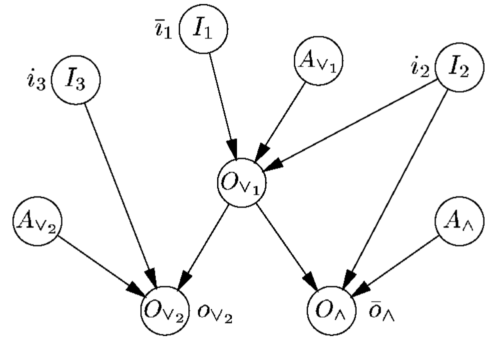

Figure 3: Bayesian diagnostic system corresponding to the circuit in Figure 1.

Definition 1 ***(Bayesian diagnostic problem)*** *A* Bayesian diagnostic problem, *denoted by $\mathcal{P}_B$, is defined as a pair $\mathcal{P}_B = (\mathcal{S}_B, \Omega)$, where $\mathcal{S}_B$ is a Bayesian diagnostic system and $\Omega = I_S \cup O_S$ denotes the set of observations.*

3.2 P-consistency and P-inconsistency

In this section, we analyse how diagnoses can be expressed in the context of Bayesian diagnostic problems.

Recall that in the consistency-based theory of diagnosis, a diagnosis is a prediction, which (*i*) assumes either normal or abnormal behaviour of each component in the system, that (*ii*) satisfies the consistency condition.

Thus, according to the first requirement, a diagnosis in a Bayesian diagnostic system concerns the entire set of behavioural assumptions. To facilitate the establishment of a connection between consistency-based diagnosis and diagnosis of a Bayesian diagnostic problem, the set of behavioural assumption for each component is denoted by Δ. By the notations a_c and $\bar{a}_c$ are meant that component c is assumed to function abnormally and normally, respectively. Clearly, Δ can then be written as

$$\Delta = \{a_c \mid c \in C\} \cup \{\bar{a}_c \mid c \in (\text{CMP} \setminus C)\}.$$

The second requirement above states that a consistency-based diagnosis has to be consistent with the observations. Note that this consistency condition implies that a diagnosis makes the observations possible. Translating this to our probabilistic diagnostic theory, the consistency condition requires that the probability of the occurrence of the observations given the diagnosis is non-zero; if this probability is equal to 0, it implies inconsistency.

These issues are embodied in the following definition.

Definition 2 ***(P-inconsistency and P-consistency)*** *Let $\mathcal{P}_B = (\mathcal{S}_B, \Omega)$ be a Bayesian diagnostic problem, then*

- *if $P(\Omega \mid \Delta) = 0$, $\mathcal{P}_B$ is called P-*inconsistent,
- *otherwise, if $P(\Omega \mid \Delta) \neq 0$, $\mathcal{P}_B$ is called P-*consistent.

The concepts of P-consistency and P-inconsistency allows us to establish a link to consistency-based diagnosis, shown in the following theorem.

Theorem 1 *Let $\mathcal{P}_L = (\mathcal{S}_L, \text{OBS})$ be a logical diagnostic problem, let $\mathcal{P}_B = (\mathcal{S}_B, \Omega)$ be a Bayesian diagnostic problem* *corresponding to this logical diagnostic problem. Let Δ be a set of behavioural assumptions. Then,*

$$P(\Omega \mid \Delta) \neq 0 \Longleftrightarrow \mathrm{SD} \cup \Delta \cup \mathrm{OBS} \nvDash \perp , \quad (4)$$

thus, the existence of a consistency-based diagnosis corresponds to P-consistency and vice versa.

Proof: [I. Flesch et al., 2006]. □

Finally, we define P-consistent diagnosis, which enables us to obtain diagnosis in a probabilistic way.

Definition 3 *(P-**consistent diagnosis**) Let $\mathcal{P}_B = (\mathcal{S}_B, \Omega)$ be a Bayesian diagnostic problem. Then, Δ is a P-*consistent diagnosis *of $\mathcal{P}_B$ iff $P(\Omega \mid \Delta) \neq 0$.*

We would like to emphasise that the notion of P-consistent diagnosis provides the basis for conflict-based diagnosis elaborated in the remainder of this paper.

4 Conflict Measure for Diagnosis

In the previous section, a Bayesian diagnostic problem was defined as a probabilistic framework that represents both qualitative and quantitative relations of a corresponding logical diagnostic problem.

The aim of this section is to show that the conflict measure can be used to distinguish between various diagnoses of a problem. In Section 4.1, we give the basic definition of the conflict measure for Bayesian diagnostic problems, which is made more specific in Section 4.2. In Section 4.3, we investigate the capability of the conflict measure to distinguish amongst various diagnoses. Finally, in Section 4.4, we derive a rational form for the conflict measure that is computationally simpler.

4.1 Basic Definition of the Conflict Measure

In this section, we define the conflict measure for Bayesian diagnostic problems, which is used as a basis for conflict-based diagnosis.

Intuitively, the conflict measure compares the probability of observing the inputs and outputs in case these observations are independent versus the case where they are dependent. If dependence between the observations is more likely than independence given a diagnosis, the conflict measure implies that there is *no* conflict. By Definition 2, however, observations of a Bayesian diagnostic problem need to be P-consistent with the problem for a given diagnosis. This implies that the definition of the conflict measure only concerns situations where the set of behavioural assumptions is equal to a diagnosis, that is, P-consistency holds. This is expressed in the following way.

Definition 4 ***(conflict measure for a Bayesian diagnostic problem)*** *Let $\mathcal{P}_B = (\mathcal{S}_B, \Omega)$ be a Bayesian diagnostic problem. Then, if $P(\Omega \mid \Delta) \neq 0$, then the conflict measure, denoted by $\mathrm{conf}_\Delta(\cdot)$, is defined as:*

$$\mathrm{conf}_\Delta(\Omega) = \log \frac{P(I_S \mid \Delta) P(O_S \mid \Delta)}{P(I_S, O_S \mid \Delta)} , \quad (5)$$

with observations $\Omega = I_S \cup O_S$.

We compute $P(I_S, O_S \mid \Delta)$ instead of $P(\Delta \mid I_S, O_S)$, as in consistency-based diagnosis, after which conflict-based diagnosis will be modelled, a diagnosis is a hypothesis that is checked. In contrast, in probabilistic abductive diagnosis one computes $P(\Delta \mid I_S, O_S)$, and Δ is a conclusion, not a hypothesis. In this paper, we are only dealing with consistency-based diagnosis.

4.2 Computation of the Conflict Measure

In this section, we derive formulae to compute the conflict measure for Bayesian diagnostic problems.

In order to derive the formulae for the conflict measure, the following assumptions are adopted. Normal behaviour is simulated in the probabilistic setting by the assumption that a normally functioning component takes an output value with probability of either 0 or 1, thus, if $\bar{a}_c$ holds, then $P(O_c \mid \pi_{O_c}) \in \{0, 1\}$. Furthermore, the set of inputs and the set of abnormality components are (marginally) independent of each other.

From now on, we assume that the inputs are conditionally independent of the output of an abnormally functioning component, i.e. $P(O_c \mid \pi_{O_c}) = P(O_c \mid a_c)$ if $a_c \in \pi_{O_c}$. We also assume that it holds that $P(o_c \mid a_c) = \alpha$, i.e., a constant probability is adopted for a given output o_c if the component c is functioning abnormally. This is a reasonable assumption in applications usually tackled by consistency-based diagnosis. Here, there is little to no knowledge of abnormal behaviour. Thus, it will be impossible to assess $P(o_c \mid a_c)$ for every component; that can be resolved by assuming them all to be equal. This approach is thus in line with previous research in consistency-based diagnosis.

As a matter of notation, $X_V = \hat{x}_V$, or simply $\hat{x}_V$, will indicate in the following that the set of random variables X_V has observed values $\hat{x}_V$. A partial assignment of values to variables X_V is written as $X_V = \tilde{x}_V$, or simply $\tilde{x}_V$, and includes observed and non-observed values.

Now, we are in the position to derive the necessary formulae for the conflict measure as given in Definition 4. The three factors of this formulae can be obtained by the application of Bayes' rule and the factorisation principle of Equation (2):

1. $P(I_S \mid \Delta) = P(I_S)$
2. $P(O_S \mid \Delta) = \sum_I P(I) \sum_{O_R} \prod_{c \in \mathrm{CMP}} P(\tilde{o}_c \mid \pi_{O_c})$
3. $P(I_S, O_S | \Delta) = \sum_{I_R} P(I) \sum_{O_R} \prod_{c \in \mathrm{CMP}} P(\tilde{o}_c | \pi_{O_c})$.

Note that in the summation $\sum_{O_R}$ we handle both observed and remaining outputs in $\tilde{o}_c$, that is, some values of the taken variables are fixed and some values are non-fixed.

Based on the equations above, we obtain

$$\mathrm{conf}_\Delta(\Omega) = \log \frac{\sum_I P(I) \sum_{O_R} \prod_{c \in \mathrm{CMP}} P(\tilde{o}_c \mid \pi_{O_c})}{\sum_{I_R} P(I_R) \sum_{O_R} \prod_{c \in \mathrm{CMP}} P(\tilde{o}_c \mid \pi_{O_c})} . \quad (6)$$

EXAMPLE 3 Consider Figure 3, which models the logical diagnostic system in Figure 1. Let $O_S = \{\bar{o}_\wedge, o_{\vee 2}\}$. The values of the conflict measure for different diagnoses and inputs are shown in Table 1. Here, we can see that the diagnosis $\{\bar{a}_\wedge, \bar{a}_{\vee_1}, \bar{a}_{\vee_2}\}$, which assumes that all components are

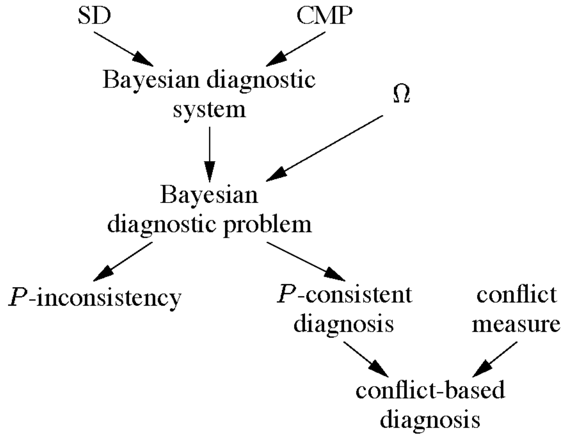

Figure 4: Framework of conflict-based diagnosis.

functioning normally, implies either P-inconsistency or P-consistency depending on the value of input I_2. Moreover, for the P-consistent diagnoses, we obtain negative, zero and positive conflict measures, indicating different relations between the observations and the patterns in the joint probability distribution. The interpretation of these results is given in Example 4.

4.3 Ordering Conflict-based Diagnoses

In this section, we show that a meaningful subset of P-consistent diagnoses can be selected based on the conflict measure; this subset of diagnoses is called conflict-based. Subsequently, we also show how the conflict measure can be used to order conflict-based diagnoses resulting in minimal conflict-based diagnoses.

We start by introducing the various Bayesian diagnostic interpretations of the conflict measure. In Example 3, it has been shown that the conflict measure can take negative, zero and positive values. Recall that in the case of a negative or zero conflict measure the joint occurrence of the observations is in accordance with the patterns in P. Therefore, in these cases, a diagnosis is called *strongly P-consistent*. A positive conflict measure is interpreted the other way around, and as it implies the existence of conflicts between the observations and P, the associated diagnoses are called *weakly P-consistent*.

The Bayesian diagnostic interpretation of the conflict measure implies that the less the value of the conflict measure the stronger P-consistent the diagnosis is, which can be understood as follows. The conflict measure favours one diagnosis above another if the behavioural assumptions of this diagnosis provide *more support* for the observed output than the behavioural assumptions of the other diagnosis. Since the probability α expresses the likelihood of positive output given abnormality, the conflict measure of a diagnosis may depend on α.

This interpretation gives rise to the following definitions and is subsequently illustrated by an example.

Definition 5 ***(conflict-based diagnosis)*** *Let $\mathcal{P}_B = (\mathcal{S}_B, \Omega)$ be a Bayesian diagnostic problem and let Δ be a P-consistent*

minimal conflict-based diagnosis $\subseteq$ conflict-based diagnosis $\subseteq$ P-consistent diagnosis

Figure 5: Set-inclusion relation between notions of diagnosis.

diagnosis of $\mathcal{P}_B$. Then, Δ is called a conflict-based diagnosis *if it is strongly P-consistent, i.e.* $\mathrm{conf}_\Delta(\Omega) \leq 0$.

Definition 6 ***(minimal conflict-based diagnosis)*** *Let Δ be a conflict-based diagnosis of $\mathcal{P}_B$. Then, Δ is called* minimal, *if for each conflict-based diagnosis Δ' it holds that* $\mathrm{conf}_\Delta(\Omega) \leq \mathrm{conf}_{\Delta'}(\Omega)$.

A summary of the framework conflict-based diagnosis is given in Figure 4, where the edges indicate how concepts are combined in defining conflict-based diagnosis.

EXAMPLE 4 We reconsider Table 1 in Example 3.

To start, we analyse the situation when $\alpha = 0.001$. Diagnosis $\Delta = \{\bar{a}_\wedge, a_{\vee_1}, a_{\vee_2}\}$ is a weakly P-consistent diagnosis for the inputs $\{i_1, i_2, i_3\}$ with value 0.0002 and a strongly P-consistent, also conflict-based, diagnosis for inputs $\{i_1, \bar{\imath}_2, i_3\}$ with -3.0002. These results can be explained as follows. For both sets of inputs, with associated outputs $\{\bar{o}_\wedge, o_{\vee_2}\}$, Δ is a possible diagnostic hypothesis according to the associated probability distribution. However, under the assumption that the AND gate is functioning normally, the input $\bar{\imath}_2$ among the inputs $\{i_1, \bar{\imath}_2, i_3\}$, offers a good match to the observed output $\bar{o}_\wedge$, whereas the match is bad for the input i_2 among the inputs $\{i_1, i_2, i_3\}$. Thus, for $\Omega = \{i_1, \bar{\imath}_2, i_3, \bar{o}_\wedge, o_{\vee_2}\}$ the diagnosis Δ is a conflict-based diagnosis, whereas it is not a conflict-based diagnosis for $\Omega' = \{i_1, i_2, i_3, \bar{o}_\wedge, o_{\vee_2}\}$.

The conflict measure for the diagnosis $\Delta' = \{a_\wedge, \bar{a}_{\vee 1}, \bar{a}_{\vee_2}\}$ remains unchanged for variation in input, because of the following reason. Since the output of the abnormally functioning AND component is observed, the conflict measure establishes which inputs are consistent with the normally functioning OR_2 and then uniformly distributes the probability over all possible inputs. There are 7 inputs that are possible of the total of 8 inputs, hence $\log 7/8 = -0.06$.

Let $\alpha = 0.001$. Then, for $\Omega = \{i_1, \bar{\imath}_2, i_3, \bar{o}_\wedge, o_{\vee_2}\}$ there are three P-consistent diagnoses mentioned in Table 1, all of them conflict-based. The minimal conflict-based diagnosis in this case is $\Delta = \{\bar{a}_\wedge, a_{\vee_1}, a_{\vee_2}\}$.

Finally, we compare conflict measure values for different values of α. When the value of α decreases from 0.5 to 0.001, $\Delta = \{\bar{a}_\wedge, a_{\vee_1}, a_{\vee_2}\}$ offers increased support for the

$\bar{o}_\wedge, o_{\vee_2}$	α	$\bar{a}_\wedge, \bar{a}_{\vee_1}, \bar{a}_{\vee_2}$	$a_\wedge, \bar{a}_{\vee_1}, \bar{a}_{\vee_2}$	$\bar{a}_\wedge, a_{\vee_1}, a_{\vee_2}$
i_1, i_2, i_3	any	–	-0.06	$\log \frac{1-1/2\alpha}{1-\alpha}$
$i_1, \bar{\imath}_2, i_3$	any	-0.43	-0.06	$\log \alpha - \frac{1}{2}\alpha^2$
i_1, i_2, i_3	0.001	–	-0.06	0.0002
$i_1, \bar{\imath}_2, i_3$	0.001	-0.43	-0.06	-0.0002
i_1, i_2, i_3	0.5	–	-0.06	0.176
$i_1, \bar{\imath}_2, i_3$	0.5	-0.43	-0.06	-0.125

Table 1: Examples of the results of the conflict measure.

observed outputs $\bar{o}_{\wedge}$ and $o_{\vee_2}$, reflected by the values 0.176 and 0.0002, and -0.426 and -3.0002. The conflict measure takes into account the likelihood of the output values of abnormally functioning components. Therefore, for the observed output $\bar{o}_{\wedge}$ more support is obtained if its inputs are negative, thus the output of component OR_1 takes on value $\bar{o}_{\vee_1}$. Since the OR_1 gate is assumed to malfunction, this has probability $1-\alpha$. But then, this probability becomes larger if the value of α decreases.

Figure 5 offers a summary of the set-inclusion relations between various notions of diagnosis defined in this paper.

4.4 A Rational Form

In this section, we show that the conflict measure can also be written in rational form, that is easier for computational purposes.

To start, we distinguish between several types of components. The sets of normally and abnormally functioning components will be denoted by $C^{\bar{a}}$ and C^{a}, respectively. These sets are separated into mutually disjoint sets of components, related to observed and remaining outputs, yielding sets $C^{\bar{a}}_S$, $C^{\bar{a}}_R$, C^{a}_S and C^{a}_R.

The relation between the abnormally functioning system components and the value α is as follows.

Lemma 1 *The joint probability distribution of the output of the abnormally functioning system components is equal to:*

$$\prod_{c\in C^a_S} P(\hat{o}_c \mid \pi_{O_c}) = \alpha^l(1-\alpha)^m\ ,$$

where l and m are the total number of components in C^a_S for which the observed output is positive or negative, respectively.

Proof: See [I. Flesch et al., 2006]. □

The conflict measure can be expressed in rational form.

Theorem 2 *Let $\mathcal{P}_B = (\mathcal{S}_B, \Omega)$ be a Bayesian diagnostic problem. Then, the conflict measure (6) is equal to:*

$$\text{conf}_\Delta(\Omega) = \log\frac{\sum_{k=0}^{q} d_k\cdot\alpha^k(1-\alpha)^{q-k}}{\sum_{k=0}^{q} e_k\cdot\alpha^k(1-\alpha)^{q-k}}\ ,$$

with $P(o_c \mid a_c) = \alpha$, $\Omega = I_S \cup O_S$, $q = |C^a_R|$ and positive constants d_k, e_k.

Proof: Consider the numerator of the conflict measure:

$$P(I_S \mid \Delta)P(O_S \mid \Delta) = \sum_I P(I)\sum_{O_R}\prod_{c\in\text{CMP}} P(\tilde{o}_c \mid \pi_{O_c})$$

$$= \sum_I P(I)\sum_{O_R}\prod_{c\in C^{\bar{a}}} P(\tilde{o}_c|\pi_{O_c})\prod_{c\in C^a_S} P(\hat{o}_c|a_c)\prod_{c\in C^a_R} P(\tilde{o}_c|a_c)$$

$$= \alpha^l(1-\alpha)^m\sum_I P(I)\sum_{O_R}\prod_{c\in C^{\bar{a}}} P(\tilde{o}_c \mid \pi_{O_c})\prod_{c\in C^a_R} P(\tilde{o}_c \mid a_c)$$

$$= \alpha^l(1-\alpha)^m\sum_I P(I)\sum_{O_R}\Psi_{O^{\bar{a}}_R}(\tilde{o}^{\bar{a}}_R, \pi_{O^{\bar{a}}_R})\cdot$$

$$\Phi_{O^{\bar{a}}_S}(\hat{o}^{\bar{a}}_S, \pi_{O^{\bar{a}}_S})\prod_{c\in C^a_R} P(\tilde{o}_c \mid a_c)$$

$$= \alpha^l(1-\alpha)^m\sum_{k=0}^{q} d_k\cdot\alpha^k(1-\alpha)^{q-k}\ ,$$

with $\Psi(\cdot,\cdot)$ and $\Phi(\cdot,\cdot)$ Boolean functions. The derivation for the denominator is similar. See [I. Flesch et al., 2006] for a full proof. □

5 Conclusions

In this paper, we proposed a new notion of model-based diagnosis, where ideas from consistency-based diagnosis and data conflict detection in statistics have been merged into one coherent framework. The result is a theory of model-based diagnosis offering features similar to those offered by consistency-based diagnosis and more, as diagnoses can be distinguished from each other using probabilistic information. We also showed that conflict-based diagnoses can be computed using a rational form.

Acknowledgements. This work has been partially funded by NWO (ProBayes project, number 612.066.201).

References

[Console et al., 1990] L. Console, D. Theseider Dupré and P. Torasso. On the relationship between abduction and deduction. *J. Logic and Computation*, 1(5):661–690, 1991.

[de Kleer et al., 1992] J. de Kleer, A. K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *AIJ*, 52:197–222, 1992.

[de Kleer, 1990] J. de Kleer. Using crude probability estimates to guide diagnosis. *AIJ*, 45:381–392, 1990.

[Genesereth, 1984] M. Genesereth. The use of design description in automated diagnosis. *AIJ*, 24:411–436, 1984.

[I. Flesch et al., 2006] I. Flesch, P.J.F. Lucas and Th.P. van der Weide. Conflict-based diagnosis: Adding uncertainty to model-based diagnosis. Technical Report 2006 (www.cs.ru.nl/~ildiko/confMeas.pdf), Radboud University Nijmegen, 2006.

[Jensen, 2001] F. V. Jensen. *Bayesian Networks and Decision Graphs*. Springer-Verlag, New York, 2001.

[Köb and Wotawa, 2004] D Köb and F. Wotawa. Introducing alias information into model-based debugging, In *Proc. DX-04*, Carcassonne, France, 2004.

[Kohlas et al., 1998] J. Kohlas, B. Anrig, R. Haenni, and P.A. Monney. Model-based diagnosis and probabilistic assumption-based reasoning. *AIJ*, 104:71–106, 1998.

[Lucas, 2001] P.J.F. Lucas. Bayesian model-based diagnosis. *Int. J. Approximate Reasoning*, 27:99–119, 2001.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems:Networks of Plausible Inference*. Morgan Kauffman, San Francisco, CA, 1988.

[Poole, 1996] D. Poole. Probabilistic conflicts in a search algorithm for estimating posterior probabilities in Bayesian networks. *AIJ*, 88(1–2):69–100, 1996.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *AIJ*, 32:57–95, 1987.

[Struss and Price, 2003] P. Struss and C. Price. Model-Based systems in the Automative industry. *AI Magazine*, 24:17–34, 2003.