



An improved diagnostic method for probabilistic consistency-based diagnosis

Marcos L. P. Bueno^{1,2*}, Arjen Hommersom^{1,3}, and Peter J. F. Lucas^{1,4}

¹ Institute for Computing and Information Sciences, Radboud University Nijmegen, the Netherlands

² CINTESIS - Centre for Health Technology and Services Research, Porto, Portugal
mbueno@cs.ru.nl

³ Faculty of Management, Science and Technology, Open University, The Netherlands
arjen.hommersom@ou.nl

⁴ Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
peter1@cs.ru.nl

Abstract

In consistency-based diagnosis (CBD), abnormal behavior is sorted out based on deviation from a normal behavior specification. Probabilities have been added to CBD for quantifying uncertainty on, e.g., the behavior of faulty components. While resulting in more complete models, the requirement of such uncertainty parameters goes in opposition to the original CBD motivation. The conflict measure stands closer to CBD by computing solutions without the need of priors on candidates, however, its results might not be suitable when only partial observations are available. In this paper, we propose a method called the *diagnostic coefficient*, which better solves the partial observability case, while needing the same parameters as the conflict measure. The diagnostic coefficient is based on the idea that observations are conflicting if the observed outputs are discrepant with respect to alternative outputs that could have been observed. We report experiments with logical circuits where the diagnostic coefficient shows promising results compared to the conflict measure under various settings with missing observations.

1 Introduction

In consistency-based diagnosis (CBD), faults are typically sorted out by contrasting observations with specifications of normal behavior [4], [12]. This means that abnormal behavior is deduced based on deviation from the nominal specification. Probabilities have been added to CBD allowing for uncertainty quantification, e.g. when specifying the behavior of a component when it is failing [2] [1], and for assigning probabilities when components are unobserved [10]. Uncertainty is also quantified when computing distributions over diagnosis candidates, e.g. as in the general diagnostic engine (GDE) [4], [3]. The disadvantage of these approaches is that they go against the primary idea in CBD of not depending on abnormal behavior specification.

*Corresponding author.

In order to reduce the dependency on parameters of probabilistic CBD, the conflict measure has been proposed [7]. The conflict measure is based on translating the system description into a Bayesian network [6], then assessing diagnoses based on the correlation between observations. The advantage of the conflict measure is that it does not require the specification of priors on diagnoses. This contrasts to, e.g., the GDE approach, which relies on abnormality priors for rendering a more abductive-like diagnostic method. While considering priors provides a more thorough approach, they are often unknown in many human-made systems, which creates the need for a more lightweight parameterization in CBD.

Research has shown a direct relationship between the standard, logic-based CBD and the conflict measure [10]. This relationship occurs when the system is fully observable, and it further aids diagnosis selection when the conflict measure is used. However, it has also showed that the interpretation of the conflict measure becomes non-trivial or inaccurate when only partial observations are available. In order to solve these issues, in this paper we propose a new diagnostic notion called *diagnostic coefficient*. The diagnostic coefficient uses the likelihood of the observations as the basis of discrepancy indicator, while not requiring specifying priors on abnormal component behavior. Furthermore, the diagnostic coefficient also takes into account the distribution of *alternative outputs* that could have been obtained under the observed inputs. This provides stronger evidence in favor or against diagnosis candidates, because it becomes possible to distinguish rare diagnoses (in the sense of poorly explanatory) from unlikely ones. This distinction becomes relevant, e.g., when observing multiple outputs.

The contributions of this paper are as follows. We first show cases where the conflict measure is not straightforward to be interpreted. Then, we introduce the diagnostic coefficient and the role of relevant alternative outputs in it. Examples are provided to illustrate the diagnostic coefficient. Finally, experiments are reported comparing the conflict measure and the diagnostic coefficient based on a real logical circuit. The experiments considered several diagnostic problems by varying the fault location, number and selection of observed and unobserved components, among other experimental parameters.

2 Preliminaries

We briefly review the necessary preliminaries on conventional and probabilistic consistency-based diagnosis.

2.1 Logic-based Consistency-based Diagnosis

In consistency-based diagnosis [4], [13], [8], normal behavior of a component c is described by logical implications of the following form:

$$\neg\text{Ab}(c) \rightarrow \text{Behavior}(c)$$

In this formalization, the literal $\neg\text{Ab}(c)$ expresses that the behavior associated with the component c only holds when the assumption that the component is not abnormal, i.e. $\neg\text{Ab}(c)$, is true for component c . For example, if component O is an OR-node in a logical circuit with two inputs (In_1 and In_2), we may write:

$$\neg\text{Ab}(O) \rightarrow (\text{In}_1(O, \text{true}) \rightarrow \text{Out}(O, \text{true}))$$

to partially specify its behavior. Logical behavior descriptions of the form discussed above are part of a *system description* SD. Classes of components can be described by quantification. In

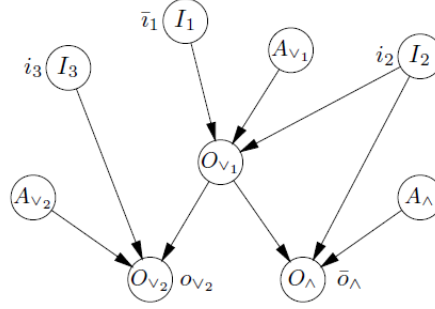


Figure 1: Bayesian diagnostic system example [7].

addition to the generic descriptions of the expected behavior of components, a system description also includes logical specifications of how the components are connected to each other (the structure of the system), and the names of the components constituting the system. Problem solving basically amounts to adopting particular assumptions about every component c , either whether $\text{Ab}(c)$ is true or false.

A diagnostic problem is then defined as a system description SD, together with a set of observations OBS, i.e., a finite set of logical formulas. Let Δ be an assignment of either a normal ($\neg\text{Ab}(c)$) or an abnormal ($\text{Ab}(c)$) behavioral assumption to each component c . Denote $\Delta_{\neg a}$ for all the normal and Δ_a for all the abnormal behavioral assumptions, i.e., $\Delta = \Delta_{\neg a} \cup \Delta_a$. We say that Δ is a *consistency-based diagnosis* iff [3]:

$$\text{SD} \cup \Delta \cup \text{OBS} \not\models \perp \quad (1)$$

Typically, in this logical approach we aim to find a subset-minimal diagnosis, i.e., a diagnosis Δ such that there is no Δ' which is also a diagnosis and $\Delta'_a \subset \Delta_a$.

2.2 Probabilistic diagnostic problems

Already from the start, uncertainty reasoning was widely recognized as an essential ingredient of a diagnostic problem solving [4]. For example, given a probabilistic model of the system, one could compute the maximum a posteriori (MAP) assignment of a set of potential diagnoses given a set of observations. Lucas [12] has proposed to combine consistency-based diagnosis and a Bayesian network approach by computing likelihoods of candidate diagnoses Δ . This can lead to a significant reduction of the number of diagnoses that have to be considered in a direct MAP approach.

In this paper, we follow the approach by Flesch and Lucas [7], which generalized consistency-based diagnosis by defining a *Bayesian diagnostic system*. A Bayesian diagnostic system is a Bayesian network (BN) with nodes $I \cup O \cup A$, where A denotes the abnormality literals. The edges of the graph are derived from a mapping from connections in SD. For more details on the mapping of SD to a Bayesian network, we refer to [6]. An example of such a Bayesian network is shown in Figure 1. Given the structure of the network, we assume that each component behaves in one of the following ways:

- **Deterministic behavior:** if a component is functioning normally, then its output is determined by a deterministic function of its inputs.

- **Random behavior:** if a component is faulted, its inputs are conditionally independent of its output. In particular, if $\pi(V)$ denotes the parents of node V in the graph, then it is assumed that for each $O_i \in O$, it holds that $P(O_i \mid \pi(O_i) \setminus \{A_i\}, a_i) = P(O_i \mid a_i)$. Moreover, we assume that $P(O_i \mid a_i)$ is fixed for every component, and we define $\gamma = P(o_i \mid a_i)$.

This assumption allows us to model intermittent faults, because a component can still output a correct value when it is faulty.

In this approach, the set of observations OBS is split into a set of input variables I and output variables O . By the notation of [7], we will denote observed inputs and outputs by I_S and O_S respectively, whereas the non-observed inputs and outputs will be denoted by I_R and O_R respectively. Thus, $I = I_S \cup I_R$ and $O = O_S \cup O_R$. Moreover, define $\Omega = I_S \cup O_S$ the full set of observations.

Interestingly, the relationship between the joint probability distribution in a Bayesian diagnostic problem and logical inconsistency is captured by the following property:

$$P(\Omega \mid \Delta) \neq 0 \text{ iff } \text{SD} \cup \Delta \cup \text{OBS} \not\models \perp$$

If $P(\Omega \mid \Delta) \neq 0$, the hypothesis Δ is called P -consistent. Thus, the existence of a consistency-based diagnosis coincides with the existence of a P -consistent diagnosis.

For P -consistent diagnoses, the situation is more subtle when using probability theories. To measure the amount of consistency, Flesch and Lucas deviate from logical consistency given by Equation 1, and instead used a probabilistic conflict measure that had been proposed to detect potential conflicts between observations and a given Bayesian network. A natural choice is then to measure the conflict between inputs and outputs in the observations, given a particular hypothesis Δ , as follows:

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

if $P(I_S, O_S \mid \Delta) \neq 0$. We will omit Ω when it is obvious in the context. In case $\text{conf}_\Delta(\Omega) \leq 0$, then the inputs and outputs are positively correlated (or uncorrelated), i.e., there is no conflict between the inputs and output. It is then said that Δ is a *conflict-based diagnosis*. A *minimal conflict-based diagnosis* is the one with the minimal conflict measure.

Although in principle abnormality priors (i.e. the distribution of the variables A in the Bayesian diagnostic system) are needed, they are not effectively used in the computation of the conflict measure. This brings the conflict measure closer to the original ideas of CBD, where typically little is known about faulty behavior.

3 Observability

3.1 Full observations

Under full observability, $\Omega = I \cup O$. For this case, properties can be identified about the relationship between logic-based CBD and the conflict measure. Proposition 3.1. shows the core computations of the conflict measure.

Proposition 3.1. *Under full observability, for any Δ consistent, it holds that:*

$$\text{conf}_\Delta = \log \sum_{\{I: \text{SD} \cup \Delta \cup I \cup O \not\models \perp\}} P(I)$$

where the summation is restricted to the inputs that are logically consistent with the remaining variables in the expression. For brevity sake, we denote by $V \not\equiv \perp$ when a set V of variables is marginalized out restricted to its values that are consistent with the rest of the expression.

Proof. Using the chain rule, the conflict measure for Δ is:

$$\text{conf}_\Delta = \log \frac{P(O \mid \Delta)}{P(O \mid I, \Delta)}$$

Using marginalization and the factorization given by the BN structure, we compute:

$$\begin{aligned} P(O \mid \Delta) &= \sum_I P(I, O \mid \Delta) \\ &= \sum_{\{I: \text{SDU} \cup \Delta \cup I \cup O \not\equiv \perp\}} P(I)P(O \mid I, \Delta) \\ &= \sum_{I \not\equiv \perp} P(I) \prod_{O_i \in O} P(O_i \mid \pi(O_i)) \end{aligned}$$

Similarly:

$$P(O \mid I, \Delta) = \prod_{O_i \in O} P(O_i \mid \pi(O_i))$$

The O_i terms can be moved outside of the summation, because they are not affected by any particular instantiation of the inputs (as long as the instantiation is consistent). This allows us to obtain the desired result. \square

The intuition for this property is that every consistent candidate is, in principle, a reasonable solution for the diagnostic instance, as the joint $P(I, O \mid \Delta)$ is at least as great as $P(I \mid \Delta) \cdot P(O \mid \Delta)$ for any such diagnosis Δ . The role of the conflict measure is in further refining the selection of which diagnosis best supports the observations. Moreover, for any consistent Δ it holds that $\text{conf}_\Delta \leq 0$. Proposition 3.1 also shows that, under full observability, the only probabilistic parameters that play a role in the conflict measure are the input priors.

Although assuming abnormalities will be often needed in order to obtain consistent diagnoses given the observed behavior, the diagnosis without abnormalities is the natural choice otherwise. This is shown in Proposition 3.2.

Proposition 3.2. *Under full observability, if the diagnosis Δ with $\Delta_a = \emptyset$ is consistent, then:*

$$\text{conf}_\Delta \leq \text{conf}_{\Delta'} \tag{2}$$

for every $\Delta' \neq \Delta$.

Proof. Suppose a diagnosis Δ' with $\Delta'_a \neq \emptyset$. By Proposition 3.1, its conflict measure will be given by the summation over consistent instantiations of all the inputs. Note that any such Δ' allows at least the same input instantiations as Δ , because adding abnormalities to components can only relax the relationship between inputs and outputs in each of them. Hence, the sum of the probabilities of joint assignments for Δ' is at least that for Δ , resulting in:

$$\text{conf}_\Delta \leq \text{conf}_{\Delta'}$$

\square

This result matches the intuition that whenever the observed outputs match the predicted outputs no component should be faulty.

3.2 Partial observations

Under partial observability, the relationship between the conflict measure and CBD becomes less straightforward. In this case, the previous properties no longer hold. The interpretation of the conflict measure is illustrated by the following examples.

Example 3.1. Consider an OR gate with two inputs and one output as shown in Figure 2. Furthermore, assume the prior $P(i_1) = 0.1$ (the same for I_2). Our assumptions tell us that $P(O_i | \pi(O_i)) = 1$ if $\neg a_i$ and inputs and output are consistent, and that $P(O_i | \pi(O_i))$ is either γ or $1 - \gamma$ if a_i (depending on the actual value observed in O_i). Thus, for the diagnosis $\Delta = \{-a_v\}$ we compute:

$$\begin{aligned} \text{conf}_\Delta &= \log \frac{P(I_S | \Delta)P(O_S | \Delta)}{P(I_S, O_S | \Delta)} \\ &= \log \frac{P(o_v | \Delta)}{P(o_v | \neg i_1, \Delta)} \\ &= \log \frac{\sum_{I_1, I_2 \neq \perp} P(I_1, I_2)P(o_v | I_1, I_2, \Delta)}{\sum_{I_2 \neq \perp} P(I_2)P(o_v | \neg i_1, I_2, \Delta)} \\ &= \log \frac{0.19}{0.1} \simeq 0.28 \end{aligned}$$

On the other hand, for the diagnosis $\Delta' = \{a_v\}$, we obtain $\text{conf}_{\Delta'} = 0$. Hence, although Δ assumes no abnormalities and is logically consistent, the diagnosis Δ' has a lower conflict than Δ . By the conflict measure, Δ' is supposed to provide a better explanation for the observed inputs and outputs. Indeed, it seems reasonable to prefer Δ' over Δ , because under Δ , consistency is achieved only when assuming I_2 to be true, which is unlikely to be the case.

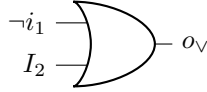


Figure 2: OR gate with partial inputs.

As Example 3.1 shows, when one has only partial observations of the system, it might be necessary to consider additional diagnoses in problem solving. Next we consider additional examples for illustrating further points in the partial observability case.

Example 3.2. Reconsider Example 3.1, however now with prior $P(i_1) = 0.9$. For the same diagnoses Δ and Δ' , we now compute $\text{conf}_\Delta \simeq 0.04$ and $\text{conf}_{\Delta'} = 0$. These results suggest that assuming the gate to be faulty should better support the observations, because Δ' lowers the conflict.

It is, however, not reasonable to make such assumption, as it is very likely that I_2 is evaluated to true, which favors the hypothesis that the gate is not faulty. This example illustrates that the diagnosis with the lowest conflict measure might not be the most intuitive one.

Example 3.3. Consider an AND gate with inputs and output as shown in Figure 3. Consider the prior $P(i_1) = 0.01$ (the same for I_2). As $I_S = \emptyset$, consistent diagnoses become indistinguishable by the conflict measure. Nevertheless, the gate is likely faulty, as the combination of inputs need to obtain the observed output is very unlikely.

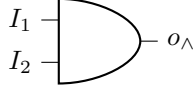


Figure 3: AND gate with missing inputs. The priors are $P(i_1) = P(i_2) = 0.01$.

3.3 Conflict measure computations

For the following analysis, we consider that O_S cannot be empty, as contrasting predicted and observed behaviors would not be possible in such situation. The denominator of the conflict measure can be written as $P(O_S | I_S, \Delta)$, and it expresses the likelihood of the observed outputs (under a diagnosis Δ and fixed inputs I_S). This term plays an important role in identifying conflicts between inputs and outputs, because the outputs are assumed to be a function of the inputs (except when the component is abnormal). Another appealing characteristic of the likelihood term is that unobserved components (e.g. unobserved inputs) can be marginalized out, so that the resultant expression will be proportional to the chances of the values of the unobserved variables that maintain consistency. Hence, input priors can be properly added into the likelihood.

The second term of the conflict measure is the marginal probability of O_S under Δ . We can see this term as a normalizing term for the likelihood of outputs. Note that both the likelihood and the normalizing term can vary under different diagnoses. Counter-intuitive behavior might arise as the sign of the conflict of different diagnoses might be shifted due to the normalizing factor, which is a term that does not distinguish seen from unseen inputs.

4 The diagnostic coefficient

4.1 Likelihood and alternative outputs

Let us denote the observed outputs by $o_b \in \text{dom}(O_S)$ and the observed inputs by $i_b \in \text{dom}(I_S)$. Then, the likelihood of o_b is given by:

$$p_0 = P(O_S = o_b | I_S = i_b, \Delta) \quad (3)$$

A low p_0 could indicate that inputs and outputs are conflicting given the corresponding Δ , and the opposite if p_0 is high. However, relying solely on the likelihood for identifying discrepancies can be uninformative: if multiple outputs are observed, it might be difficult to distinguish discrepant probabilities, because the probability mass must be distributed across many possible outputs.

In order to make the identification of discrepant outputs more robust, we consider the distribution of *alternative* outputs that could have been observed. This means that we investigate $P(O_S | i_b, \Delta)$ for values that O_S can take on other than o_b . By analyzing the location of p_0 in such distribution we are able to gather more solid evidence for deciding on the discrepancy of the observations. To this end, let us define the following quantity:

$$p_{\text{alt}} = \sum_{\substack{o_{\text{alt}} \in \text{dom}(O_S) \\ o_{\text{alt}} \neq o_b}} \left\{ P(o_{\text{alt}} | i_b, \Delta) : P(o_{\text{alt}} | i_b, \Delta) \leq p_0 \right\} \quad (4)$$

which considers the probability mass of the alternative outputs that are consistent with i_b , and that are equally or less likely than o_b .

By inducing a partitioning of the probability mass of O_S , p_{alt} can help one determine whether o_b is rare (in the sense of out of ordinary, discrepant) when it is unlikely. Given a set of observations Ω , we combine p_0 and p_{alt} in order to define the *diagnostic coefficient* as:

$$\text{dc}_\Delta(\Omega) = \max \{p_0, p_{\text{alt}}\} \quad (5)$$

where dc takes values on $[0, 1]$. A related notion to p_{alt} is the surprise index [9], [14], which has been used for example in model revision [11].

In order to select a diagnosis from a set of candidates, we first select those that are likely not out-of-ordinary diagnoses. We implement this notion by defining a diagnosis Δ as *rejected* if $\text{dc}_\Delta \leq \alpha$, where α is a real number in $[0, 1]$; otherwise, we say that Δ is *accepted* (in this work, $\alpha = 0.05$). The rejected diagnoses are only discarded if there is an accepted diagnosis among all the candidates. The final step ranks the remaining diagnoses according to their p_0 and takes the one with the highest p_0 ; in case of ties, the smaller one (in terms of number of abnormalities) is preferred.

4.2 Examples

Example 4.1. Reconsider Example 3.2. Suppose $\gamma = 0.5$. For the diagnosis $\Delta = \{\neg a_v\}$, we compute:

$$p_0 = P(o_v \mid \neg i_1, \neg a_v) = \sum_{I_2 \neq \perp} P(I_2) = 0.9$$

$$p_{\text{alt}} = P(\neg o_v \mid \neg i_1, \neg a_v) = 0.1$$

Hence, $\text{dc}_\Delta = 0.9$. On the other hand, for $\Delta' = \{a_v\}$ we compute:

$$p_0 = P(o_v \mid \neg i_1, a_v) = P(o_v \mid a_v) = \gamma = 0.5$$

$$p_{\text{alt}} = 0.5$$

Thus, $\text{dc}_{\Delta'} = 0.5$. Based on the diagnosis selection criteria previously stated, none of the diagnoses are rejected, and we choose Δ . The results provided by the diagnostic coefficient match the intuition that O_v is likely not faulty.

Example 4.2. Reconsider Example 3.3 and the same diagnoses as before. With Δ , we obtain $p_0 \simeq 0$ and $p_{\text{alt}} = 0$, as the alternative output $\neg o_\wedge$ has a higher likelihood than o_\wedge . Thus, $\text{dc}_\Delta \simeq 0$. For the diagnosis Δ' , it holds that $p_0 = p_{\text{alt}} = 0.5$, thus $\text{dc}_{\Delta'} = 0.5$.

Based on the selection criteria, Δ is rejected and Δ' is accepted, hence Δ' is the selected diagnosis. Selecting Δ' is indeed reasonable, as previously discussed in Example 3.3.

4.3 Full and partial observations

We investigate next a few properties of the diagnostic coefficient. We start off with the diagnosis that assumes abnormalities on every component (also known as the trivial diagnosis).

Proposition 4.1. *Let Δ be a diagnosis such that $\Delta = \Delta_a$. Then, it holds that:*

$$p_0 = \gamma^u (1 - \gamma)^v$$

$$p_{\text{alt}} = \sum_{\substack{o_{\text{alt}} \in \text{dom}(O_S) \\ o_{\text{alt}} \neq o_b}} \{ \gamma^{u_{\text{alt}}} (1 - \gamma)^{v_{\text{alt}}} : \gamma^{u_{\text{alt}}} (1 - \gamma)^{v_{\text{alt}}} \leq p_0 \}$$

where u and v denote the number of true and false values in o_b respectively. Similarly, u_{alt} and v_{alt} count such occurrences in each alternative o_{alt} .

Proof. As all components are assumed to be faulty, the outputs are independent of the inputs:

$$p_0 = P(o_b \mid i_b, \Delta) = P(o_b \mid \Delta) = \gamma^u(1 - \gamma)^v$$

For p_{alt} , each alternative o_{alt} will assign values to the outputs different than o_b . Thus, the counts will differ from those in p_0 , and we obtain:

$$P(o_{\text{alt}} \mid i_b, \Delta) = \gamma^{u_{\text{alt}}}(1 - \gamma)^{v_{\text{alt}}}$$

which allows us to obtain the desired result. \square

With the conflict measure, $\text{conf}_\Delta = 0$ for the trivial diagnosis regardless of the observations, as shown in [10]. By opposition, the diagnostic coefficient for the trivial diagnosis depends on the value of the parameter γ . Although the trivial diagnosis might not be often suitable, it seems reasonable that the diagnostic coefficient be sensitive to γ , which allows the method to assess whether the observations are likely to be seen under such diagnosis.

Proposition 4.2. *For any consistent diagnosis Δ , it holds that:*

$$p_0 = \sum_{I_R, O_R \neq \perp} P(I_R) \gamma^u (1 - \gamma)^v$$

$$p_{\text{alt}} = \sum_{\substack{o_{\text{alt}} \in \text{dom}(O_S) \\ o_{\text{alt}} \neq o_b}} \{P(o_{\text{alt}} \mid i_b, \Delta) : P(o_{\text{alt}} \mid i_b, \Delta) \leq p_0\}$$

where

$$P(o_{\text{alt}} \mid i_b, \Delta) = \sum_{I_R, O_R \neq \perp} P(I_R) \gamma^{u_{\text{alt}}} (1 - \gamma)^{v_{\text{alt}}}$$

Here, u and v denote the number of true and false values in the faulty components of $O = O_S \cup O_R$, respectively. In other words, this considers o_b as well as the consistent instantiations for O_R .

Proposition 4.2 shows how the computations in the diagnostic coefficient are split into unobserved inputs (if there is any) and random outputs from the faulty components.

Proposition 4.3. *Under full observations, for any consistent diagnosis Δ it holds that:*

$$p_0 = \gamma^u(1 - \gamma)^v$$

$$p_{\text{alt}} = m \cdot \gamma^u(1 - \gamma)^v$$

where m denotes the number of consistent alternative outputs (given i_b and Δ).

Proposition 4.3 is relevant because it shows that the probabilistic parameters of component failure are effectively used in the diagnostic coefficient. While it seems natural to include such parameters both in the full and in the partial observability cases, in the conflict measure they appear only in the latter case. Secondly, as opposed to the conflict measure, the diagnostic coefficient form in Proposition 4.3 does not depend on priors of inputs, which seems intuitive here because all the components are observable.

As a corollary of Proposition 4.3, note that if Δ is the diagnosis with no abnormality assumptions and it is consistent under full observations, then $\text{dc}_\Delta = 1$. This implies that such Δ becomes the natural choice in that case, because for any other diagnosis Δ' it holds that $\text{dc}_{\Delta'} \leq \text{dc}_\Delta$.

5 Experimental results

In this section we report experiments involving the conflict measure and the diagnostic coefficient. The implementation is based on ProbLog (for logical and probabilistic reasoning) [5] and Python¹.

5.1 Diagnostic instances

In order to evaluate the diagnostic methods, we consider the full adder circuit. It is composed by 3 inputs and 5 outputs, as shown in Figure 4. Each diagnostic instance is created by first assigning inputs (fixed for all instances) to the circuit and deducing the outputs by logical reasoning (given no abnormalities). Next, a fault is injected into one output by switching its value. Then, we select which input and output variables will be observed for composing the observation set, where we require that at least one output is observed.

A broad range of diagnostic problems is covered in the experiments: all the possible single fault placings, as well as all choices of inputs and outputs are considered. The number of observations range from 3 to 8 components, where the latter means full observability. In this setting, therefore, the *true* diagnosis of each instance is known beforehand.

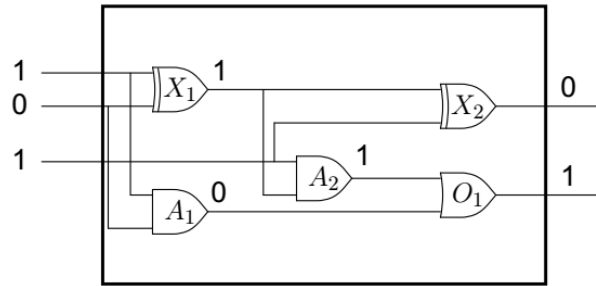


Figure 4: Full adder with observations.

5.2 Evaluation criteria

For the conflict measure, diagnoses are ranked based on their conflict measure value; the diagnosis (or diagnoses, in case of ties) with the lowest conflict and minimal number of abnormalities is selected. For the diagnostic coefficient, the ranking is obtained as dictated in Section 4.1. In order to assess the output provided by the diagnostic methods, we define two criteria as follows:

- **Hit rate.** Let us consider the best-ranked diagnoses that are tied in a given diagnostic instance. We register whether the true diagnosis belongs to such collection of diagnoses. The hit rate indicates the average of such indicator over a set of diagnostic instances.
- **Weighted hit rate.** Suppose the true diagnosis is included in the top (tied) ones in a given diagnostic instance. What is the chance that we pick the true diagnosis in such collection? This is provided by the weighted hit rate, considering a set of diagnostic instances. The weighted hit rate corresponds to the hit rate weighted by the mean probability that the true diagnosis is selected.

¹Available at <http://www.cs.ru.nl/~mbueno/>

5.3 Results

The results for the full and partial observability cases based on the full adder are shown in Figure 5. Under full observability, the results suggest that both the conflict measure and the new diagnostic coefficient detected the faulty components on all the cases. The conclusions are, however, different when only partial observations are available (i.e. when $N_{obs} \leq 7$): in general, reducing the number of observations tends to decrease the accuracy of both diagnostic methods. The effect of that on each method is, however, not the same.

The hit rate values suggest that the diagnostic coefficient identifies the true diagnosis much more often than the conflict measure. In other words, the true diagnosis has indeed the best value in dc very often, which is certainly not the case with conf. On the other hand, the ties in dc are on average larger than those in conf, leading to greater reduction in the weighted hit rate compared to the reduction experienced in this indicator by conf. On average, the results suggest that the diagnostic ability of dc is more accurate than that of conf.

Inst.	Hit rate(%)		W. Hit rate(%)	
	conf	dc	conf	dc
D1	100	100	100	100
D2	100	100	100	100
D3	100	100	100	100
D4	100	100	100	100
D5	100	100	100	100

(a) $N_{obs} = 8$.
Diagnostic instances = 1.

Inst.	Hit rate(%)		W. Hit rate(%)	
	conf	dc	conf	dc
D1	46.4	75*	46.4	61.9*
D2	57.1	75*	55.4	66.1*
D3	75	89.3*	58.3	67.3*
D4	60.7	75*	53	61.9*
D5	60.7	75*	51.2	63.7*

(c) $N_{obs} = 6$.
Diagnostic instances = 28.

Inst.	Hit rate(%)		W. Hit rate(%)	
	conf	dc	conf	dc
D1	17.1	52.9*	11.7	38.3*
D2	35.7	45.7*	28.1	40*
D3	45.7*	41.4	33.6*	23.6
D4	32.9*	28.6	25.1*	18.6
D5	31.4	37.1*	25	27*

(e) $N_{obs} = 4$.
Diagnostic instances = 70.

Inst.	Hit rate(%)		W. Hit rate(%)	
	conf	dc	conf	dc
D1	75	87.5*	75	81.2*
D2	75	87.5*	75	81.2*
D3	87.5	100*	75	87.5*
D4	75	87.5*	68.8	81.2*
D5	75	87.5*	68.8	81.2*

(b) $N_{obs} = 7$.
Diagnostic instances = 8.

Inst.	Hit rate(%)		W. Hit rate(%)	
	conf	dc	conf	dc
D1	26.8	64.3*	24.1	47.3*
D2	44.6	62.5*	40.2	52.7*
D3	62.5	67.9*	46.4*	44.6
D4	50	57.1*	42.1*	40.8
D5	48.2	58.9*	38.4	45.8*

(d) $N_{obs} = 5$.
Diagnostic instances = 56.

Inst.	Hit rate(%)		W. Hit rate(%)	
	conf	dc	conf	dc
D1	9.1	38.2*	4.8	29.1*
D2	30.9	36.4*	21.8	34.5*
D3	25.5*	16.4	19.7*	8.5
D4	18.2*	10.9	9.8*	6.4
D5	20*	16.4	16.1*	12.1

(f) $N_{obs} = 3$.
Diagnostic instances = 55.

Figure 5: Results on the full adder circuit. Diagnostic instances refer to the number of ways inputs and outputs can be chosen for composing the observations. Hence, each row averages over such instances, given a fixed faulty component. The best results are shown in bold face followed by an asterisk.

5.4 Discussion

The difficulty faced by both the diagnostic methods seems reasonable: the fewer the observations, the weaker the diagnostic capability of the method. The essential components for accurately identifying the faulty component might not belong to the observation set. This happens, e.g., if components affected by the injected fault (beyond the faulty component itself) are not observed. A limitation of the experiments here reported is that we did not distinguish *diagnosable* from *non-diagnosable* instances, hence, the reported performance of the methods is likely underestimated. By selecting a proper subset of diagnostic instances, a more insightful and optimistic assessment of diagnostic methods (and still realistic) could be obtained.

6 Conclusions

In this paper we investigated model-based diagnosis, in particular consistency-based diagnosis, in the context of full and partial observability. We discussed difficult cases that might arise with existing approaches to the probabilistic formulation of CBD, specially when diagnosing under partial observations. Motivated by that we proposed a new diagnostic notion called diagnostic coefficient. The empirical results comparing these methods showed that the diagnostic coefficient was more accurate on the average and is more intuitive to be interpreted.

As future work, we plan investigating other possible ways to combine the likelihood of observations with the likelihood of alternatives, which are the central ideas of the diagnostic coefficient. Furthermore, in order to use the proposed method on larger problems, a deeper analysis of the computational cost of the diagnostic coefficient is needed. This might lead to approximations that should still be able to capture the most relevant alternative observations that help identify faults. Finally, we would like to evaluate multiple fault cases, which we believe could be suitably solved by the proposed method as well.

Acknowledgments

This work has been funded by NWO (Netherlands Organisation for Scientific Research, project Careful) and NORTE 2020 (project NanoSTIMA). Project “NORTE-01-0145-FEDER-000016” (NanoSTIMA) is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

We also thank the anonymous reviewers for their valuable remarks.

References

- [1] Rui Abreu and Arjan J.C. van Gemund. Diagnosing multiple intermittent failures using maximum likelihood estimation. *Artificial Intelligence*, 174(18):1481 – 1497, 2010.
- [2] Johan De Kleer. Diagnosing multiple persistent and intermittent faults. In *Proc. of the 21st Int. Joint Conf. on Artificial Intelligence*, IJCAI’09, pages 733–738, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.
- [3] Johan De Kleer, Alan K Mackworth, and Raymond Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56(2-3):197–222, 1992.
- [4] Johan de Kleer and Brian C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1):97 – 130, 1987.

- [5] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI'07*, pages 2468–2473, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.
- [6] Ildikó Flesch. *On the use of independence relations in Bayesian networks*. PhD thesis, Radboud University Nijmegen, 2008.
- [7] Ildikó Flesch, Peter J. F. Lucas, and Theo P. van der Weide. Conflict-based diagnosis: Adding uncertainty to model-based diagnosis. In *IJCAI 2007, Proc. of the 20th Int. Joint Conf. on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, page 380, 2007.
- [8] K.D. Forbus and J. De Kleer. *Building Problem Solvers*. The MIT Press, Cambridge, MA, 1993.
- [9] J. Habbema. Models for diagnosis and detection of combinations of diseases. In *Decision Making and Medical Care*, pages 399–411, 1976.
- [10] Arjen Hommersom and Marcos L. P. Bueno. Toward computing conflict-based diagnoses in probabilistic logic programming. In *Proc. of the 3rd Int. Workshop on Probabilistic Logic Programming co-located with 26th ILP, London, UK, September 3, 2016*, pages 29–38, 2016.
- [11] Kathryn Blackmond Laskey. Conflict and surprise: Heuristics for model revision. In *Proc. of the Seventh Conference on Uncertainty in Artificial Intelligence, UAI'91*, pages 197–204, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.
- [12] Peter J.F. Lucas. Bayesian model-based diagnosis. *International Journal of Approximate Reasoning*, 27(2):99–119, 2001.
- [13] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57 – 95, 1987.
- [14] Adam Zagorecki, Marcin Kozniewski, and Marek Druzdzel. An approximation of surprise index as a measure of confidence. In *AAAI Fall Symposium Series*, pages 39–41, 2015.