Toward Computing Conflict-Based Diagnoses in Probabilistic Logic Programming

Arjen Hommersom[1,2] and Marcos L.P. Bueno[2]

[1] Open University of the Netherlands
[2] Radboud University Nijmegen, The Netherlands

Abstract. Consistency-based diagnosis is a well-known theory of diagnosis using main knowledge of the normal structure and behaviour of a system. Central in this theory is the notion of a *conflict*, which describes the relationship between the knowledge of the system and observations. In literature, at least two notions of a conflict have been proposed: (1) based on logical inconsistencies of the observations w.r.t. the system and (2) based on a probabilistic conflict measure. Probabilistic logic programming languages combine probabilities with logic, so this raises the question to which extent these two notions coincide. In this paper, we consider consistency-based diagnosis in ProbLog and discuss some preliminary results on the relationship between logical consistency and the probabilistic conflict measure.

1 Introduction

Model-based diagnosis concerns itself with identifying faults based on a description (a model) of the system [7]. With the ever increasing models that are learned from data, diagnostic methods are still very applicable, see e.g. a recent overview of model-based diagnosis in medicine [10].

Most of the foundational work on model-based diagnosis was done in the 1980s, in particular in consistency-based diagnosis [6,13,4] and abductive diagnosis [11]. Certainly in consistency-based diagnosis there will be many candidate hypotheses that explain the observations. To make a choice between these candidates, the probability of a diagnosis has been proposed as a way to rank the different hypothesis. Early work in this area includes the general diagnostic engine (GDE) [6], which computes probabilities of candidate diagnosis by making strong independence assumptions. In more recent years, Bayesian networks together with a probabilistic *conflict measure* has been proposed as an alternative [3].

An advantage of the logical approach is that the system description can be written down in a first-order logical language which provides a concise theory of the system. In recent years several probabilistic logical programming (PLP) languages have been proposed that can be seen as relational generalisations of Bayesian networks, therefore forming an attractive foundation for probabilistic model-based diagnosis. There is obviously a close relationship between PLP and

abduction reasoning, which was recognised in early papers (in particular [12]), but the relationship to consistency-based diagnosis approach is less clear.

In this paper, we discuss a preliminary exploration of conflict-diagnosis for diagnostic problems modelled in ProbLog [1] programs. We show that abductive explanations of probabilistic logic programs can be exploited to find conflict-based diagnoses.

2 Preliminaries

We briefly review the necessary preliminaries on consistency-based and conflict-diagnosis, given both logical models and in a probabilistic setting.

2.1 Logic-based Consistency-based Diagnosis

The following description is based on a recent survey on clinical model-based diagnosis [10]. As described in this survey, in consistency-based diagnosis, in contrast to abductive diagnosis, the malfunctioning of a system is diagnosed by using mainly knowledge of the normal structure and normal behaviour of its components [6,13,4].

In this paper, we are concerned with what is called Deviation-from-Normal-Structure-and-Behaviour diagnosis. For this type of diagnosis, we consider knowledge concerning normal structure and behaviour of the system. In contrast, little or no explicit knowledge is available about the relationships between abnormalities, on the one hand, and findings to be observed when certain disorders are present, on the other hand. From a practical point of view, the primary motivation for investigating this approach to diagnosis is that in many domains little knowledge concerning abnormality is available, which is certainly true for new human-created artifacts. For example, during the development of a new printer, experience with respect to the faults that may occur when the printing system is in operation is lacking. Thus, the only conceivable way in which initially such faults can be handled is by looking at the normal structure and functional behaviour of the printer.

In consistency-based diagnosis, normal behaviour of a component c is described by logical implications of the following form:

$$\neg\mathrm{Ab}(c) \rightarrow \mathrm{Behaviour}(c)$$

In this formalisation, the literal $\neg\mathrm{Ab}(c)$ expresses that the behaviour associated with the component c only holds when the assumption that the component is not abnormal, i.e. $\neg\mathrm{Ab}(c)$, 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:

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

to partially specify its behaviour. Logical behaviour descriptions of the form discussed above are part of a *system description* SD. Classes of components can
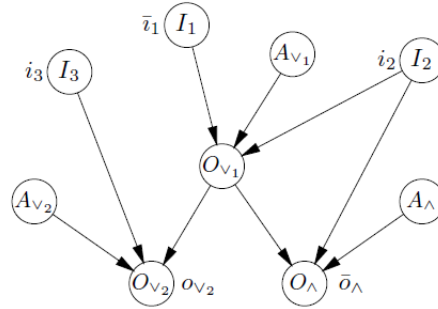
Fig. 1. Example Bayesian diagnostic system from [3].

be described by quantification. In addition to the generic descriptions of the
expected behaviour 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. Prob-
lem solving basically amounts to adopting particular assumptions about every
component c, either whether $\mathrm{Ab}(c)$ 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 ($\neg\mathrm{Ab}(c)$) or an abnormal ($\mathrm{Ab}(c)$) behavioural
assumption to each component c. Denote $\Delta_{\neg a}$ for all the normal and Δ_a for all
the abnormal behavioural assumptions, i.e., $\Delta = \Delta_{\neg a} \cup \Delta_a$. We say that Δ is a
consistency-based diagnosis iff [8]:

$$\mathrm{SD} \cup \Delta \cup \mathrm{OBS} \not\models \bot \tag{1}$$

Typically, in this logical approach we aim to find a subset-minimal diagnosis,
i.e., a diagnosis Δ such that there does not exists a 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 recognised as an essen-
tial ingredient of a diagnostic problem solving [6]. For example, given a proba-
bilistic model of the system, one could compute the maximum a posterior (MAP)
assignment of a set of potential diagnosis given a set of observations. Lucas [9]
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 [3], which gen-
eralised consistency-based diagnosis to *conflict-based diagnosis*. They define a

Bayesian diagnostic system as a Bayesian network 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 [2]. An example of such a Bayesian network is shown in Fig. 1. Given the structure of the network, it is assumed that the inputs are conditionally independent of the output of a particular component if the component is functioning abnormally. 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, it is assumed that the distribution of $P(O_i \mid a_i)$ is fixed of every component. Furthermore, if the component is functioning normally, then the output of the component is determined by a deterministic function of its inputs.

In this approach, the set of observations OBS is split into a set of inputs variables I and output variables O. By the notation of [3], we will denote observed inputs and outputs by I_S and O_S, whereas the remaining non-observed outputs are denoted by I_R and O_R, i.e., $I = I_S \cup I_R$ and $O = O_S \cup O_R$. Furthermore, 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:

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

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. It is defined as follows [5].

$$\mathrm{conf}(\Omega) = \log \frac{P(\Omega_1)P(\Omega_2)\cdots P(\Omega_n)}{P(\Omega)}$$

A natural choice is then to measure the conflict between the input and output in the observations, given a particular hypothesis Δ, as follows:

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

if $P(I_S, O_S \mid \Delta) \neq 0$. In case $\mathrm{conf}_\Delta(\Omega) \leq 0$, 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.

3 Consistency-based diagnosis in ProbLog

Given a system description SD, a diagnostic logic programming theory is simply defined as ProbLog program that encodes a Bayesian diagnostic system. Each
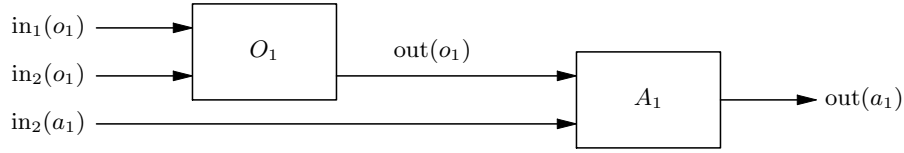
Fig. 2. Simple logical circuit with an OR-gate O_1 and AND-gate A_1.

random variable is mapped to a literal, such that each input is represented by `in`n`(COMP, Value)`, where n is the nth input of the component, the output represented by `out(COMP, Value)`, and each abnormality random variable is represented by `ab(COMP)`. Without loss of generality, `Value` is assumed to have groundings `true` and `false`. Furthermore, by abuse of notation, we will refer to these literals by I, O, and A as there is a one-to-one mapping of these literals to random variables in the diagnostic Bayesian network.

Using annotated disjunctions [14], a diagnostic PLP $\mathcal{L}$ contains the following clauses:

- α`::ab(C).`
- β`::in(C,true); `$(1-\beta)$`::in(C,false).`
- γ`::out(C, true); `$(1\text{-}\gamma)$`::out(C,false) :- ab(C).`
- `out(C,Value) :- \+ ab(C), b(inputs).`
  where `inputs` is a set of literals consisting of inputs of C `in`k`(C,I`k`)` or outputs from other components `out(C',V)`, and b is a Boolean function.

Given the assumptions mentioned in Section 2.2, and fixing the prior distribution of inputs and abnormalities, it is straightforward to see that the distribution of a Bayesian diagnostic system and this PLP system is the same.

Example 1. Consider a simple circuit as depicted in Fig. 2. The system can be encoded by a diagnostic PLP as follows:

```
0.1::ab(C).
0.5::in1(C,true) ; 0.5::in1(C,false).
0.5::in2(C,true) ; 0.5::in2(C,false).

0.5::out(C, true) ; 0.5::out(C,false) :- ab(C).

out(o1, true)  :- \+ ab(o1), (in1(o1, true) ; in2(o1,true)).
out(o1, false) :- \+ ab(o1), in1(o1,false), in2(o1, false).

out(a1, true)  :- \+ ab(a1), in1(a1, true), out(o1, true).
out(a1, false) :- \+ ab(a1), (in1(a1, false) ; out(o1,false)).
```

One apparent advantage of using a PLP is that we can specify the local logical structure directly, whereas in a Bayesian network, logical formulas have to be encoded into the CPT. Furthermore, a significant advantage is that we can

benefit from logical variables to specify the behaviour of classes of component, e.g. OR-components.

In the remainder of this section, we consider some properties of these diagnostic logic programs. If we are using PLP we are both in a logical and a probabilistic setting, so the natural question is: what is the relationship between consistency-based diagnosis and conflict-based diagnosis in a diagnostic PLP?

However first, we note the following about conflict-based diagnoses.

Proposition 1. *For any set of observations given a diagnostic PLP, there is a conflict-based diagnosis.*

Proof. Observe that there is a conflict-based diagnosis in a trivial sense, namely, if $\Delta = \Delta_a = \{a \mid a \in A\}$, i.e., when all components are abnormal. In this case all outputs are conditionally independent of the inputs given Δ, and therefore:

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

$$= \log \frac{P(I \mid \Delta)P(O \mid \Delta)}{P(O \mid \Delta)P(I \mid \Delta)} = 0 \leq 0$$

3.1 Complete observation

We first consider the case where we have complete observations on the inputs and outputs, i.e., $\Omega = I \cup O$. The following propositions show that in this complete case consistency and conflict-based diagnosis coincide.

Proposition 2. *Let Ω be a complete assignment with inputs and outputs. If* $\mathrm{SD} \cup \Omega \cup \Delta \not\models \bot$, *then:*

$$\mathrm{conf}_\Delta(\Omega) \leq 0$$

i.e., Δ is a conflict-based diagnosis.

Proof. Let $a_i \in A$ be an abnormality literal associated to the output $O_i \in O$. It then holds that $P(O_i \mid \pi(O_i)) = 1$ if $\neg a_i \in \pi(O_i)$ and $P(O_i \mid \pi(O_i)) = P(O_i \mid a_i)$ otherwise.

Given these observations, the conditional probability of the output given the input can be written as follows:

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

i.e., it is only determined by the constant γ and not by the value of the inputs. Then, by similar reasoning, we have:

$$P(O \mid \Delta) = \sum_I P(O, I \mid \Delta) = \sum_I P(O \mid I, \Delta)P(I)$$

$$= \sum_{\{I \mid \mathrm{SD} \cup \{O, I\} \cup \Delta \not\models \bot\}} \prod_{a_i \in \Delta_a} P(O_i \mid a_i)P(I)$$

$$= \prod_{a_i \in \Delta_a} P(O_i \mid a_i) \sum_{\{I \mid \mathrm{SD} \cup \{O, I\} \cup \Delta \not\models \bot\}} P(I)$$

Therefore $P(O \mid I, \Delta) \geq P(O \mid \Delta)$.

It then follows:

$$\mathrm{conf}_\Delta(\Omega) = \log \frac{P(I \mid \Delta)P(O \mid \Delta)}{P(I, O \mid \Delta)} = \log \frac{P(O \mid \Delta)}{P(O \mid I, \Delta)} \leq 0$$

This is relevant, because logic programming can be used to detect this consistency in a direct way, shown in the following proposition.

Proposition 3. *Let $\mathcal{L}$ be a diagnostic logic program. There is an explanation E consisting of choices for each probabilistic fact in $\mathcal{L}$ that proves Ω, iff* $\mathrm{SD} \cup \mathrm{OBS} \cup \Delta \not\models \bot$*, where $\Delta \subseteq E$.*

Proof. ($\Leftarrow$) Directly by the construction of the PLP program. ($\Rightarrow$) We have $\mathcal{L} \cup E \models \Omega$, hence, the assignment on the abnormality literals ensures that the inputs and outputs are consistent.

Propositions 2 and 3 are convenient, because this means that for complete observations, we can always obtain a conflict-based diagnosis by means of *abductive reasoning alone*, i.e., without computing marginal probabilities from the PLP.

Example 2. In the remainder of the paper, we use a very simple diagnostic PLP with only a single OR-gate, specified as follows.

```
0.01::ab(C).
0.5::in1(C,true);0.5::in1(C,false).
0.5::in2(C,true);0.5::in2(C,false).

0.5::out(C, true);0.5::out(C,false) :- ab(C).

out(o1, true)  :- \+ ab(o1), (in1(o1, true) ; in2(o1,true)).
out(o1, false) :- \+ ab(o1), in1(o1,false), in2(o1, false).
```

Suppose $\Omega = \{\mathtt{in1(o1, false)}, \mathtt{in2(o1, false)}, \mathtt{out(o1, true)}\}$, and $\Delta = \{\mathtt{\backslash+\ ab(o1)}\}$. Obviously the only explanation for Ω assumes that O_1 is abnormal, hence Δ is not a consistency-based nor a conflict-based diagnosis. Also, by Proposition 3, this explanation shows that $\Delta' = \{\mathtt{ab(o1)}\}$ is a consistency-based diagnosis. Note that by Proposition 1, this is also a conflict-based diagnosis, in particular $\mathrm{conf}_{\Delta'}(\Omega) = 0$.

3.2 Partial observations

Now suppose partial observability, i.e., if $\Omega \subsetneq I \cup O$. The following example shows that when Δ is P-consistent, then conflict-based diagnosis and consistency-based diagnosis does *not necessarily* coincide.

Example 3. Reconsider the diagnostic program of Example 2. Given this program, suppose we take $\Omega = \{\texttt{in1(o1, false)}, \texttt{out(o1, true)}\}$, and $\Delta = \{\texttt{\textbackslash+ ab(o1)}\}$, which is clearly a consistency-based diagnosis, we have:

$$P(I \mid \Delta) = 0.5$$
$$P(O \mid \Delta) = 0.75$$
$$P(I, O \mid \Delta) = 0.25$$

and therefore:

$$\text{conf}_\Delta(\Omega) = \log \frac{0.5 \cdot 0.75}{0.25} \simeq 0.18$$

This example shows that logical reasoning cannot be used to derive a conflict-based diagnosis alone, something that would be expected. Of course, in this case, the only conflict-based diagnosis is again the trivial one by Proposition 1, namely $\Delta' = \{\texttt{ab(o1)}\}$, because then $\text{conf}_{\Delta'}(\Omega) = 0$.

The question is now: given a consistency-based diagnosis, can we extend this to a non-trivial conflict-based diagnosis? The following proposition suggests that abductive explanations can also be used for this.

Proposition 4. *Let $\mathcal{L}$ be a diagnostic PLP and Δ be a hypothesis. If $\Delta'_a = \Delta_a \cup a'$, with a' being an abnormality predicate for one of the observed outputs O' such that $\text{conf}_{\Delta'}(\Omega) < \text{conf}_\Delta(\Omega)$, then there is an explanation for $T = \{\mathcal{L} \cup \Delta_a\}$ that proves $I_S \cup \{\neg O_i \mid O_i \in O_S\}$ and contains $\neg a'$.*

Proof. (sketch) We first repeatedly apply the chain rule on the conflict measure given Δ' by the structure of the underlying Bayesian network:

$$\text{conf}_{\Delta'}(\Omega) = \log \frac{P(O_S \mid \Delta')}{P(O_S \mid I_S, \Delta')}$$
$$= \log \frac{P(O' \mid \Delta', \pi(O') \cap O_S)}{P(O' \mid I_S, \Delta', \pi(O') \cap O_S)} \prod_{O \in O_S \setminus \{O'\}} \frac{P(O \mid \Delta', \pi(O) \cap O_S)}{P(O \mid I_S, \Delta', \pi(O) \cap O_S)}$$

Note that $P(O' \mid \Delta', \pi(O') \cap O_S) = P(O' \mid I_S, \Delta', \pi(O') \cap O_S) = P(O' \mid a')$, so the first term in the product is equal to 1.

Now by contraposition, suppose all explanations for $I_S \cup \{\neg O_i \mid O_i \in O_S\}$ given T contain a'. Then for the associated output O' it holds $T \cup I_S \cup \neg O'$ implies a', or equivalently: $\mathcal{L} \cup \Delta_a \cup \neg a' \cup I_S \models O'$. This then implies that $P(O' \mid I_S, \pi(O') \cap O_S, \Delta) = 1$, and therefore:

$$\frac{P(O' \mid \pi(O') \cap O_S, \Delta)}{P(O' \mid I_S, \pi(O') \cap O_S, \Delta)} = P(O' \mid \pi(O') \cap O_S, \Delta) \leq 1$$

Additionally, for all $O \in O_S \setminus \{O'\}$, we have $P(O \mid \Delta', \pi(O) \cap O_S) = P(O \mid \Delta, \pi(O) \cap O_S)$ and $P(O \mid I_S, \Delta', \pi(O) \cap O_S) = P(O \mid I_S, \Delta, \pi(O) \cap O_S)$. Thus:

$$\text{conf}_{\Delta'}(\Omega) \geq \text{conf}_\Delta(\Omega)$$
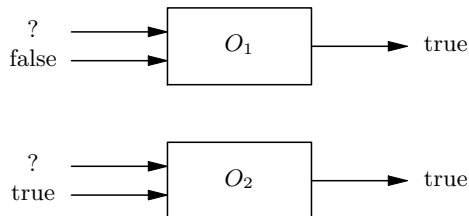
which concludes the proof.

Fig. 3. Circuit with observations used in Example 4.

The intuition is that explanations for alternative outputs for the same input reduces the correlation between the observed input and output. Hence, these components are more likely to act abnormally, so we conjecture that the converse of the property above also holds.

Example 4. Consider the problem of Example 3, and add another OR-node:

```
out(o2, true)  :- \+ ab(o2), (in1(o2,true) ; in2(o2,true)).
out(o2, false) :- \+ ab(o2), in1(o2,false), in2(o2,false).
```

Consider $\Omega = I_S \cup O_S$ with:

$$I_S = \{\texttt{in2(o1,false)}, \texttt{in2(o2,true)}\}$$
$$O_S = \{\texttt{out(o1,true)}, \texttt{out(o2,true)}\}$$

This situation is depicted in Fig. 3. If $\Delta = \{\texttt{\textbackslash+ ab(o1)}, \texttt{\textbackslash+ ab(o2)}\}$, then $\text{conf}_\Delta(\Omega) \simeq 0.05$. The most likely explanation which includes I_S but not O_S that contains a normality assumption is:

$$\texttt{in2(o1,false)}, \texttt{in2(o2,true)}, \texttt{\textbackslash+ ab(o1)}, \texttt{in1(o1,false)}$$

i.e., the observed input for O_1 could have produced a different output, suggesting that the observed input and output of O_1 are not strongly correlated. Indeed, if we consider $\Delta' = \{\texttt{ab(o1)}, \texttt{\textbackslash+ ab(o2)}\}$, we obtain $\text{conf}_{\Delta'} = -0.12$, whereas if $\Delta'' = \{\texttt{\textbackslash+ ab(o1)}, \texttt{ab(o2)}\}$, we compute $\text{conf}_{\Delta''} = 0.18$.

4 Conclusions

In this paper, we have argued for using PLP as a basis for model-based diagnosis. Furthermore, we show how consistency-based and conflict-based diagnosis can be formalised as part of a single framework. In case there are complete observations, the two notions coincide, whereas if certain inputs or outputs remain unobserved, there is a qualitative difference between the two notions.

Furthermore, we have provided some preliminary ideas on how to compute conflict-based diagnosis from abductive explanations. A naive approach for computing a conflict-based diagnosis would need to consider all possible hypotheses, which is exponential in the number of components. Instead, the observations we made in this paper suggests that computing conflict-based diagnoses can be guided by explanations computed from a given diagnostic probabilistic logic program.

Acknowledgements

References

1. Fierens, D., Van den Broeck, G., Renkens, J., Shterionov, D., Gutmann, B., Thon, I., Janssens, G., De Raedt, L.: Inference and learning in probabilistic logic programs using weighted boolean formulas. Theory and Practice of Logic Programming 15(03), 358–401 (2015)
2. Flesch, I.: On the use of independence relations in Bayesian networks. Ph.D. thesis, Radboud University Nijmegen (2008)
3. Flesch, I., Lucas, P.J., van der Weide, T.P.: Conflict-based diagnosis: Adding uncertainty to model-based diagnosis. In: IJCAI. vol. 2007, pp. 380–385 (2007)
4. Forbus, K., De Kleer, J.: Building Problem Solvers. The MIT Press, Cambridge, MA (1993)
5. Jensen, F., Nielsen, T.: Bayesian networks and decision graphs. Springer Science & Business Media (2009)
6. de Kleer, J., Williams, B.: Diagnosing multiple faults. Artificial Intelligence 32, 97–130 (1987)
7. Kleer, J.d.: Local methods for localizing faults in electronic circuits. Tech. rep., DTIC Document (1976)
8. Kleer, J.d., Mackworth, A.K., Reiter, R.: Characterizing diagnoses and systems. Artificial Intelligence 56(2-3), 197–222 (1992)
9. Lucas, P.J.: Bayesian model-based diagnosis. International Journal of Approximate Reasoning 27(2), 99–119 (2001)
10. Lucas, P.J., Orihuela-Espina, F.: Representing knowledge for clinical diagnostic reasoning. In: Foundations of Biomedical Knowledge Representation, pp. 35–45. Springer (2015)
11. Poole, D.: Normality and faults in logic-based diagnosis. In: IJCAI. vol. 89, pp. 1304–1310 (1989)
12. Poole, D.: Probabilistic Horn abduction and Bayesian networks. Artificial intelligence 64(1), 81–129 (1993)
13. Reiter, R.: A theory of diagnosis from first principles. Artificial Intelligence 32, 57–95 (1987)
14. Vennekens, J., Verbaeten, S., Bruynooghe, M.: Logic programs with annotated disjunctions. In: International Conference on Logic Programming. pp. 431–445. Springer (2004)