



HAL
open science

Using consistency and abduction based indices in possibilistic causal diagnosis

Olivier de Mouzon, Didier Dubois, Henri Prade

► **To cite this version:**

Olivier de Mouzon, Didier Dubois, Henri Prade. Using consistency and abduction based indices in possibilistic causal diagnosis. 9th IEEE International Conference on Fuzzy Systems (FUZZ- IEEE 2000), May 2000, San Antonio (Texas), United States. pp.729-734, 10.1109/FUZZY.2000.839122 . hal-03403394

HAL Id: hal-03403394

<https://hal.science/hal-03403394>

Submitted on 27 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using Consistency and Abduction Based Indices in Possibilistic Causal Diagnosis*

O. De Mouzon, D. Dubois and H. Prade

IRIT

Université Paul Sabatier
118, route de Narbonne
F-31062 Toulouse cedex 04
FRANCE

Abstract—Causal diagnosis deals with the search for plausible causes which may have produced observed effects. Knowledge about possible effects of a malfunction on a given attribute is represented by a possibility distribution, as well as the possible values of an observed attribute (giving the imprecision of the observation). Any kind of attributes (binary, numerical, ...) is allowed. In this paper, we shall restrict to single-fault diagnosis. Two main indices, respectively based on consistency and on abduction, enable us to discriminate the malfunctions. The case where one deals with imprecise information only is first discussed and exemplified. The extension to information pervaded with uncertainty is then studied. Refinements of indices are also considered.

I INTRODUCTION

Different fuzzy set-based approaches to diagnosis have been proposed for the past 25 years. On the one hand, some try to identify abnormal situations (e.g. [10]). On the other hand, some seek for possible reasons that could explain the observations (as in [11], [8] and [1]).

This paper is concerned with the second problem.

Sanchez ([11]) first suggested the use of fuzzy relation equations in order to express that faults have effects through causal links. In this type of approach, fuzziness can model:

- an intensity degree expressing to what extent a proposition is satisfied (for instance, a *quite* strong fever, where *quite* refers to some specific membership level for the fuzzy predicate "strong");
- an uncertainty (or certainty) degree attached to the observation, or to the presence of an effect, when a specific malfunction occurs.

The approach developed in [2] is based on the latter interpretation for malfunctions involving binary attributes. It introduces indices that evaluate to what extent some (possibly uncertain) observations discard (or, on the contrary, suggest)

*This work takes place in a joint project between IRIT and Siemens Automotive S.A. on a diagnosis system for car engine tuning.

the presence of malfunctions (the effects of which are not always known with certainty). A formal generalization of this approach to non-binary attributes was recently proposed in [5].

The given indices will be discussed in the general case, in this paper, under the single-fault hypothesis. An example of their use is given and some refinements are also considered. Section II presents the problem under concern and the chosen approach. The particular case where the information is only imprecise is first studied and exemplified in Section III. The general case, where the information may also be pervaded with uncertainty, is considered and illustrated in Section IV.

II MAIN ISSUES AND NOTATIONS

The main goal of this joint project is to be able to identify malfunctions (faults) given some expert knowledge and some observations (read on sensors). As we can imagine, the representation of observations is quite simple (binary, discrete and numerical attributes), even though the (im)precision of sensors needs to be modeled. One difficulty of this work is basically due to experts' knowledge representation, which yields a more or less complex and reliable knowledge base for the diagnosis. This knowledge is of the causal type, but mainly relies on previous experience (in particular, no statistics over the malfunctions are available). Fuzzy sets are used to represent the expert knowledge and the observations. The paper describes several indices and ranks them according to their capability of providing the best diagnosis process. The case of crisp sets is first considered before being extended to fuzzy sets. The paper shows the advantages of fuzzy sets compared to crisp sets.

Let \mathcal{M} be the set of all (known) possible malfunctions and \mathcal{A} be the set of the n observable attributes: $\{X_1, \dots, X_n\}$. Let $m \in \mathcal{M}$ and $i \in \{1, \dots, n\}$, then π_m^i is the possibility distribution which gives the (more or less) plausible values for attribute X_i when malfunction m (only¹) is present. Let U_i be the domain of X_i , so $\pi_m^i : U_i \rightarrow [0, 1]$. \mathcal{K}_m^i will be the fuzzy set corresponding to possibility distribution π_m^i . It represents what is known about the effects of malfunction m on attribute

¹The whole paper assumes the single-fault hypothesis.

X_i . \mathcal{K}_m^i is also called *effect*, or *symptom*, of m on X_i (for instance, if m is present, an effect is that the temperature is at least 38°C).

The observations may also be imprecise (or uncertain). $\mu_{\mathcal{O}_i} : U_i \rightarrow [0, 1]$ is the possibility distribution, which gives the (more or less) plausible values for the observed value of attribute X_i . \mathcal{O}_i denotes the fuzzy set corresponding to possibility distribution $\mu_{\mathcal{O}_i}$. It expresses the imprecision (or uncertainty) of the observations (coming from sensors). In other words, it represents the possible actual values for attribute X_i .

\mathcal{K}_m^i and \mathcal{O}_i both express imprecision, when they contain more than one element. Yet, they give information of two highly different types:

- imprecision for \mathcal{O}_i can be “controlled”: we could get more precise (but may be more expensive) or less precise information by changing the sensors;
- imprecision on \mathcal{K}_m^i , on the contrary, cannot be reduced (or changed) that easily: it depends on the available knowledge about the system only.

Note that when attribute X_i is not yet observed, its value is not known, and it could be any value of U_i : $\forall u \in U_i, \mu_{\mathcal{O}_i}(u) = 1$. Similarly, when malfunction m has no known effect on attribute X_i , all values are allowed: $\forall u \in U_i, \pi_m^i(u) = 1$. In fact, we could as well not bother specifying π_m^i explicitly: In the following, when no information is given on a π_m^i , it will implicitly mean that $\forall u \in U_i, \pi_m^i(u) = 1$.

III IMPRECISE INFORMATION CASE

In this part, only crisp sets are considered:

- $\forall m \in \mathcal{M}, \forall i \in \{1, \dots, n\}, \forall u \in U_i, \pi_m^i(u) \in \{0, 1\}$;
- $\forall i \in \{1, \dots, n\}, \forall u \in U_i, \mu_{\mathcal{O}_i}(u) \in \{0, 1\}$.

\mathcal{K}_m^i is here the (crisp) set of all possible values for attribute X_i when malfunction m is present: $\mathcal{K}_m^i = \{u \in U_i \mid \pi_m^i(u) = 1\}$. $\bar{\mathcal{K}}_m^i$ is the complement of \mathcal{K}_m^i in U_i : $\bar{\mathcal{K}}_m^i = \{u \in U_i, \pi_m^i(u) = 0\}$. Thus, $\bar{\mathcal{K}}_m^i$ contains the impossible values for attribute X_i when m is present.

This means that for each malfunction, we asked experts to give us very precise bounds, for each attribute, between the values they considered possible and the ones they considered impossible (when the given malfunction occurs). Some fictitious (but realistic and typical) information we could get from the experts are represented in Table I, where we have six malfunctions (one of them could be “air inlet temperature out of order”), $\mathcal{M} = \{m_1, m_2, m_3, m_4, m_5, m_6\}$, and four attributes: $X_1 = T$ is a temperature (numeric), $X_2 = P$ is a pressure (numeric too), $X_3 = KD$ represents knock detection (binary: Yes or No) and $X_4 = AFR$ is the air/fuel ratio (discrete: Poor, Stoechniometric or Rich). No graph is provided (π_m^i not defined) when experts know nothing about the effects

of m_i on X_i (equivalent to $\forall u \in U_i, \pi_m^i(u) = 1$). On the contrary, if $\exists u \in U_i, \pi_m^i(u) = 0$, then m has an effect on X_i : it restricts the possible values for X_i to $\mathcal{K}_m^i \subsetneq U_i$ (and we also have $\bar{\mathcal{K}}_m^i \neq \emptyset$). The more values are forbidden, the more observations may discard m . And if \mathcal{K}_m^i contains only one element of U_i , m has a precise effect on X_i (with respect to U_i granularity). $\mathcal{O}_i = \{u \in U_i \mid \mu_{\mathcal{O}_i}(u) = 1\}$ is the (crisp) set of all possible actual values (which are compatible with the observation, o_i measured on the sensor) for attribute X_i . This mainly depends on the sensor reliability. In the whole paper, we will take as example a $-10\%/+15\%$ error tolerance on o_1 and a ± 3 unit tolerance on o_2 . As for all binary attributes, X_3 is observed with perfect precision ($X_3 = \{o_3\} = \{Y\}$ or $\{N\}$). For X_4 , it depends on the air/fuel ratio measured, o_4 : We know that, according to the sensor reliability, if o_4 is under 0.95 then the ratio is poor ($\mathcal{O}_4 = \{P\}$). If $o_4 \in [0.95, 0.98]$ then the ratio might be poor or stoichiometric ($\mathcal{O}_4 = \{P, S\}$). Similarly, $o_4 \in [0.98, 1.03] \Rightarrow \mathcal{O}_4 = \{S\}$, $o_4 \in [1.03, 1.07] \Rightarrow \mathcal{O}_4 = \{S, R\}$, $o_4 > 1.07 \Rightarrow \mathcal{O}_4 = \{R\}$ (the ratio is rich).

The diagnosis process is based on the confrontation between the known effects of the system and the observations. The question is thus to know if an effect of malfunction m on attribute X_i is present or not.

There are three main cases (and answers):

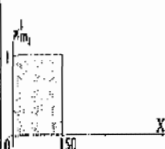
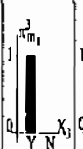

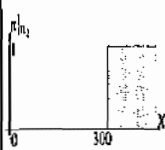
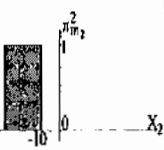

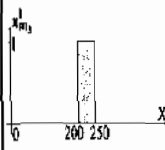
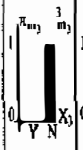

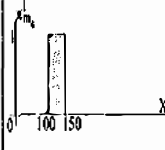
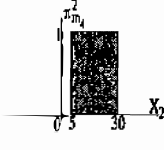
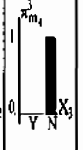
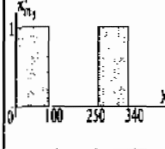
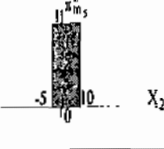


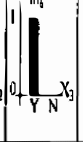
- $\mathcal{K}_m^i \cap \mathcal{O}_i = \emptyset$: the effect is absent;
- $\mathcal{K}_m^i \cap \mathcal{O}_i \neq \emptyset$: the effect might be present (or absent);
- a special case of the latter is $\mathcal{O}_i \subseteq \mathcal{K}_m^i$: the effect is present (it is sure that the effect of malfunction m on attribute X_i is present). m is (still) a valid candidate, all the more as the effect of m on X_i is precise (i.e. \mathcal{K}_m^i has a “small” size). Note that this does not mean that m is present: this particular effect of m is present, but its other effects could be absent and the observed effect could come from another malfunction m' (in this case, m and m' would both have this effect).

Another particular case is the reverse inclusion: $\mathcal{K}_m^i \subsetneq \mathcal{O}_i$ (i.e. $\mathcal{K}_m^i \cap \bar{\mathcal{O}}_i = \emptyset$, with $\bar{\mathcal{O}}_i = \{u \in U_i, \mu_{\mathcal{O}_i}(u) = 0\}$, the complement of \mathcal{O}_i in U_i). This test is positive if and only if all the effects of a malfunction may have been observed. It does not tell us whether an effect is certainly present or not. Besides, in our study, the \mathcal{O}_i are thinner than the \mathcal{K}_m^i so the test will be negative whatever the observations. We will thus not use this inclusion.

We then define two sets corresponding to the $\mathcal{K}_m^i \cap \mathcal{O}_i$ and the $\mathcal{O}_i \subseteq \mathcal{K}_m^i$ tests for all attributes:

- the set of malfunctions *all the effects of which could be present* (they all are consistent with the observations): $\text{CONS} = \{m \in \mathcal{M} \mid \forall i \in \{1, \dots, n\}, \mathcal{K}_m^i \cap \mathcal{O}_i \neq \emptyset\}$;

TABLE I
CRISP EFFECTS OF EACH MALFUNCTION ATTRIBUTE PER ATTRIBUTE

m_j	X_1 ($^{\circ}C$)	X_2 ($10^2 Pa$)	X_3 (Y/N)	X_4 (P/S/R)
m_1		a		
m_2			a	
m_3		a		
m_4				a
m_5			a	
m_6	a			a

a No graph is provided ($\pi_{m_j}^i$ not defined) when experts know nothing about the effects of m_j on X_i (equivalent to $\forall u \in U_i, \pi_{m_j}^i(u) = 1$).

- the set of malfunctions all the effects of which are sure to be present (they all are relevant to the observations): $REL = \{m \in \mathcal{M} \mid \forall i \in \{1, \dots, n\}, \mathcal{O}_i \subseteq \mathcal{K}_m^i\}$.

When one of these sets contains a given malfunction, it has a strong meaning, because all the attributes have successfully passed the consistency or relevance test.

A careful approach to diagnosis would be based on $CONS$ only, which gives all the possibly present malfunctions. For instance, suppose we measured $o_1 = 200^{\circ}C$, $o_2 = 0 Pa$, $o_3 = N$ and $o_4 = 1.05$. We get $\mathcal{O}_1 = [180, 220]$, $\mathcal{O}_2 = [-3, 3]$, $\mathcal{O}_3 = \{N\}$ and $\mathcal{O}_4 = \{S, R\}$. So $CONS = \{m_3\}$ (only m_3 and m_6 are left by X_1 , m_2, m_4 and m_6 are rejected by X_2 , m_1 and m_6 are rejected by X_3 and X_4 rejects no malfunction). So m_3 may be seriously suspected to be present. Yet, $REL = \emptyset$ (X_1 could be under $200^{\circ}C$)...

N.B.: the influence of the computation order for the different attributes (in order to optimize the diagnosis response time) will not be studied in this paper.

Similarly, $CONS = \emptyset$ means that no malfunction could be suspected (with respect to the knowledge formalized in the \mathcal{K}_m^i). When $o_1 = 274^{\circ}C$, $o_2 = 10 Pa$, $o_3 = Y$ and $o_4 = 0.96$. We get $\mathcal{O}_1 = [246.8, 301.4]$, $\mathcal{O}_2 = [7, 13]$, $\mathcal{O}_3 = \{Y\}$ and $\mathcal{O}_4 = \{P, S\}$. So $CONS = \emptyset$ (X_1 rejects m_1 and m_4 , m_2 , then m_2 and m_6 are rejected by X_2 , m_3 and m_4 are rejected by X_3 and X_4 rejects m_5). So m_3 may be seriously suspected to be present.

Yet, $CONS$ may select so many malfunctions that it gives no real help to identify a malfunction. In these cases, REL (subset of $CONS$) can be useful as it will list the most likely malfunctions only. For example, if $o_1 = 100^{\circ}C$, $o_2 = 20 Pa$, $o_3 = Y$ and $o_4 = 0.96$, then $\mathcal{O}_1 = [90, 110]$, $\mathcal{O}_2 = [17, 23]$, $\mathcal{O}_3 = \{Y\}$ and $\mathcal{O}_4 = \{P, S\}$. We then get $CONS = \{m_1, m_6\}$. And $REL = \{m_1\}$ will tell us which malfunction to suspect first (m_6 is rejected because X_2 could be under $20 Pa$).

As we have seen, REL may discard malfunctions which could be present. This is why we will first have to use the $CONS$ index. Yet, when $CONS$ is "too large" ($card(CONS) > 1$), REL may help refine the diagnosis and select a more specific malfunction (all the effect of which must have been observed). We show here that REL has abductive characteristics.

Note that when the observations are precise (i.e. $card(\mathcal{O}_i) = 1$), $\bullet_i \subseteq \mathcal{K}_m^i \iff \mathcal{K}_m^i \cap \mathcal{O}_i \neq \emptyset$ holds (i.e. $CONS = REL$). If all the observations are precise, it follows that $REL = CONS$. On the contrary, if the observations are too imprecise, $REL = \emptyset$!

Indeed, if $o_1 = 274^{\circ}C$, $o_2 = -7.5 Pa$, $o_3 = N$ and $o_4 = 1.05$, then $\mathcal{O}_1 = [246.8, 301.4]$, $\mathcal{O}_2 = [-10.5, -4.5]$, $\mathcal{O}_3 = \{N\}$ and $\mathcal{O}_4 = \{S, R\}$. We then get $CONS = \{m_2, m_3, m_5\}$. But $REL = \emptyset$ (no possible malfunction is selected). So we still do not know which malfunction should be chosen first.

TABLE II
FREQUENCIES OF THE DIFFERENT DIAGNOSIS RESPONSE

Set \ Cardinal	0	1	2	3	> 3
CONS	56.1%	39.3%	4.6%	0.0008%	0%
REL			1.3%	0%	
REL'			3.2%	0.0008%	

For this reason, we need to improve *REL*. A possible way for this is to relax *REL* in *REL'* selecting malfunctions most of the effects of which have certainly be observed among the malfunctions still remaining after *CONS*.

With the example presented just above, we get: $REL'(m) = 1$, $REL'(m_3) = 3$, $REL'(m_5) = 1$, so we can chose m_3 first.

Let rel' be:
 $rel' : \mathcal{M} \rightarrow \{0, \dots, n\}$
 $m \mapsto 0$, if $m \notin CONS$
 $m \mapsto \sum_{i=1}^n Incl(m, i)$, if $m \in CONS$

and $Incl : \mathcal{M} \times \{1, \dots, n\} \rightarrow \{0, 1\}$
 $(m, i) \mapsto 0$, if $\mathcal{O}_i \not\subseteq \mathcal{K}_m^i$
 $(m, i) \mapsto 1$, if $\mathcal{O}_i \subseteq \mathcal{K}_m^i$.

Let $REL' = \{m \in \mathcal{M} \mid rel'(m) = \max_{m' \in \mathcal{M}} (rel'(m'))\}$. Note that *REL* selects the m such that $rel'(m) = n$. *REL'* is thus at least as good as *REL* (when $REL \neq \emptyset$, $REL' = REL$) and is much better in some other cases (when $REL = \emptyset$). In fact we have: $REL \subseteq REL' \subseteq CONS \subseteq \mathcal{M}$. So, *REL'* can be useful only when $card(CONS) > 1$ and $REL = \emptyset$.

For instance, when $REL = CONS$ and $card(CONS) > 1$, we also have $REL' = CONS$, which is not sufficient to discriminate among the selected malfunctions. In fact, as soon as $card(REL') > 1$ we need other refinements to single out a specific malfunction. For this purpose, we will study in future work the specificity of an effect: the more \mathcal{K}_m^i is "small" (i.e. the effect of m_i on X_i is rather precise), the more informative it is to observe this effect.

In order to have an idea on the frequencies of the different diagnosis responses, we have improved a counting method presented in [4]. The results concerning some frequencies on this section example (restricted to bounded measured values) are presented in Table II. The chosen bounds are $[0, 500]$ for o_1 , $[-20, 40]$ for o_2 and $[0.8, 1.2]$ for o_4 . Table II shows that *CONS* is efficient in most of the cases (95.4% of the value domain: $[0, 500] \times [-20, 40] \times \{Y, N\} \times [0.8, 1.2]$). It also shows that *REL'* is much more efficient than *REL*: *REL'* solves 68.8% of the cases where $card(CONS) = 2$ whereas *REL* solves 27.6% of them only; besides *REL'* solves all the cases where $card(CONS) = 3$ and *REL* solves none of them. Note that all this does not take into account that some values are more frequently measured than others.

IV UNCERTAIN INFORMATION CASE

In fact, for the knowledge representation, the experts would be very often more comfortable in expressing their confidence in

some values they consider highly possible, some others they consider totally impossible and for all the other values, they are not really sure about what to say. For continuous attributes, the experts only need to tell what they know best (values of possibility 0 and 1) and π_m^i is then computed to follow the given information and to be continuous and piecewise linear (see example below). For discrete attributes, we can have different levels of possibility (in the example – see below – of this section: from 0 to 1 by step of 0.1 units).

This is why sets become fuzzy in this part:

- $\forall m \in \mathcal{M}, \forall i \in \{1, \dots, n\}, \forall u \in U_i, \pi_m^i(u) \in [0, 1]$;
- $\forall i \in \{1, \dots, n\}, \forall u \in U_i, \mu_{\mathcal{O}_i}(u) \in [0, 1]$.

And thus, the uncertain information case is more general than the imprecise one.

Taking the same example as in the previous section, but defining more precisely the knowledge of the experts together with them, we could get the following knowledge base (Table III). For instance, concerning malfunction m_1 and its effects on the temperature (X_1), the experts will finally say that values over $150^\circ C$ are totally impossible (when m_1 occurs) and that they expect values rather under $100^\circ C$.

The observed values are still classical (crisp) sets because they come from sensors which only have error tolerances.

A A Consistency-Based Index

In this approach, we define a consistency-based index: $\mu_{cons} : \mathcal{M} \rightarrow [0, 1]$, which enables to discard observation-inconsistent malfunctions. In fact, μ_{cons} is the fuzzy counterpart of *CONS*: malfunction m is all the more discarded as $\mu_{cons}(m)$ is close to 0.

The possibility distribution attached to $\mathcal{O}_i \cap \mathcal{K}_m^i$ is defined by: $u \mapsto \min(\mu_{\mathcal{O}_i}(u), \pi_m^i(u))$. The elements of highest possibility in this intersection give the consistency degree between \mathcal{O}_i and \mathcal{K}_m^i : $\sup_{u \in U_i} \min(\mu_{\mathcal{O}_i}(u), \pi_m^i(u))$. The consistency degree for any malfunction m with the observations is then given according to those of \mathcal{O}_i and \mathcal{K}_m^i (for each attribute):

$$\mu_{cons}(m) = \min_{i=1}^n \sup_{u \in U_i} \min(\mu_{\mathcal{O}_i}(u), \pi_m^i(u)). \quad (1)$$

The previous crisp solution might look more simple. So we could instead try to get back to it either by taking into account the core of \mathcal{K}_m^i only (all values with possibility degree of 1) or its whole support (all values with non 0 possibility degree). Let's call the corresponding consistency sets $CONS_{core}$ and $CONS_{support}$ (respectively). Note that $CONS_{core} \subseteq CONS_{support}$ holds.

Yet, $CONS_{core}$ might be empty (when a malfunction should be suspected) and $CONS_{support}$ might select too many malfunctions. For instance, if $o_1 = 200^\circ C$, $o_2 = 7.5 Pa$, $o_3 = N$ and $o_4 = 1$. We get $\mathcal{O}_1 = [180, 220]$, $\mathcal{O}_2 = [4.5, 10.5]$, $\mathcal{O}_3 = \{N\}$ and $\mathcal{O}_4 = \{S\}$. So $CONS_{core} = \emptyset$ and $CONS_{support} = \{m_3, m_4, m_5\}$. Which m should be preferred?

TABLE III
FUZZY EFFECTS OF EACH MALFUNCTION ATTRIBUTE PER ATTRIBUTE

$\pi_{m_j}^1$	X_1 ($^{\circ}C$)	X_2 ($10^2 Pa$)	X_3 (Y/N)	X_4 (P/S/R)
m_1		b		
m_2			b	
m_3		b		
m_4				b
m_5			b	
m_6	b			b

^b No graph is provided ($\pi_{m_j}^i$ not defined) when experts know nothing about the effects of m_j on X_i (equivalent to $\forall u \in U_i, \pi_{m_j}^i(u) = 1$).

Note that $CONS_{core} = \{m \in \mathcal{M} \mid \mu_{cons}(m) = 1\}$ and $CONS_{support} = \{m \in \mathcal{M} \mid \mu_{cons}(m) > 0\}$.

This is why the consistency-based index in the fuzzy approach is better than a consistency set from the crisp approach. Besides, [5] shows how this expression is close to the Bayesian approach, under ("natural") conditions.

Thus, the use of fuzzy sets allows for a more accurate representation of the knowledge and then the fuzzy set $CONS_{fuzzy}$ (defined by index μ_{cons}) rank-orders the malfunctions according to their plausibility.

Yet, in case of too incomplete knowledge, μ_{cons} might not be sufficient in order to select a small enough number of malfunctions, especially when $CONS_{core}$ selects more than one malfunction: When $o_1 = 230^{\circ}C$, $o_2 = 3Pa$, $o_3 = N$ and $o_4 = 1.1$. We get $\mathcal{O}_1 = \{207, 253\}$, $\mathcal{O}_2 = \{0, 6\}$, $\mathcal{O}_3 = \{N\}$ and $\mathcal{O}_4 = \{R\}$. So $\mu_{cons}(m_3) = \mu_{cons}(m_5) = 1$ and the other malfunctions have $\mu_{cons}(m_i) < 1$. Then, $\mu_{rel}(m_3) = 0.94$ (reached with $X_1 = 253^{\circ}C$) and $\mu_{rel}(m_5) = 0.14$ (0.8 for $X_2 = 6Pa$ and 0.14 for $X_1 = 207^{\circ}C$). Thus, we shall first suspect m_3 .

Finally, a second indice is required in order to refine μ_{cons} and bring a better conclusion: find, among the undiscarded malfunctions, which one is most relevant to the observations.

For that purpose, we used an abductive approach.

B Abduction-Based Indices

In this part, we would like to single out observation-connected malfunctions. The idea is to use the previously defined REL , REL' or equivalent. Yet, the fuzzy aspect of the data require to define a fuzzy inclusion. This will define an abductive index, as shown below.

Let us first consider the case of a single attribute ($n = 1$), say X . Let \mathcal{O} be the only observation. Extension of the results to several attributes ($n \geq 2$) is then very easy.

In classical binary logic, abduction amounts to suggesting a malfunction hypothesis, m , which could explain observation \mathcal{O} , according to some knowledge ($m \Rightarrow \mathcal{K}_m$). In classical logic, identity $\mathcal{O} = \mathcal{K}_m$ holds and abduction can be written: $\frac{\mathcal{K}_m, m \Rightarrow \mathcal{K}_m}{\mathcal{O} = \mathcal{K}_m}$.

Here, the situation is more complex:

- the observation is not precise (\mathcal{O} is a disjunctive fuzzy set);
- the knowledge is not precise (when malfunction m is present, then attribute X is a disjunctive fuzzy set \mathcal{K}_m);
- $\mathcal{O} \neq \mathcal{K}_m$.

The link between " \mathcal{O} " and " $m \Rightarrow \mathcal{K}_m$ " in order to suggest m has now to be made through fuzzy sets \mathcal{O} and \mathcal{K}_m . This link will define a fuzzy abduction.

In [5], two indices are given for this purpose. They represent, in fact, two particular cases corresponding to fuzzy inclusions of \mathcal{K}_m in \mathcal{O} , for one of them, and \mathcal{O} in \mathcal{K}_m , for the other.

For crisp sets, $A \subset B$ is equivalent to $\forall x, x \in A \Rightarrow x \in B$. The fuzzy inclusion degree can thus be seen as the lowest

degree of $A(x) \rightarrow_f B(x)$, for all x , where $A(x)$ and $B(x)$ are the (respective) membership degrees of x to fuzzy sets A and B and where \rightarrow_f is a fuzzy implication, which can be chosen among some already known fuzzy implications.

In the following, only fuzzy inclusion of \mathcal{O} in \mathcal{K}_m will be studied: as in the crisp case, it is the only one that guarantees (for an inclusion degree of 1) the presence of symptom \mathcal{K}_m of m . We can then use the abductive scheme.

In the general case, where there are n attributes, the weakest fuzzy inclusion tells the extent to which the malfunction is likely and we will note μ_{rel} the corresponding index defined (for any m) by:

$$\mu_{rel}(m) = \min_{i=1}^n \inf_{u \in U} \mu_{\mathcal{O}_i}(u) \rightarrow_f \pi_m^i(u). \quad (2)$$

$$\mu_{rel}(m) = 1 \iff \forall i \in \{1, \dots, n\}, \mathcal{O}_i \subseteq \mathcal{K}_m^i.$$

Now, an appropriate fuzzy implication (\rightarrow_f) has to be chosen. For this purpose, we studied Gödel's² and Goguen's³ residuated implications, Dienes'⁴ and Reichenbach's⁵ strong implications, and Łukasiewicz's implication⁶ (both strong and residuated).

In this paper, a strong implication will be chosen because values 1 (classical implication) are more significant with these implications ($\forall i \in \{1, \dots, n\}$, the support of \mathcal{O}_i needs to be included in the core of \mathcal{K}_m^i). Besides, we (always) have $\rightarrow_D \leq \rightarrow_R$. So Dienes' implication is even more discriminating. This is why this implication is chosen.

μ_{cons} and μ_{rel} are then completely defined and we may draw a parallel with their crisp versions, *CONS* and *REL*:

- $REL \subseteq CONS$: $\mu_{rel} \leq \mu_{cons}$ (provided that π_m^i is normalised, which is true here: every malfunction has at least one totally possible value for each attribute);
- whenever observations are precise, $REL = CONS$: here also, $\mu_{rel} = \mu_{cons}$ (intersections $\mathcal{O}_i \cap \mathcal{K}_m^i$ are singletons, where the only element u has a possibility degree of $\pi_m^i(u)$).

Finally, the diagnosis is first based on μ_{cons} in order to rank the malfunctions and then on μ_{rel} in case of twin first malfunctions. Reference [7] discusses such indices in another application. Note that it is much more difficult to get twins in $[0, 1]$ than in $\{0, 1\}$. Should there still be some twin malfunctions, we could define a μ_{rel}' with a sum, a leximin or a leximax depending on the required behaviour (compensatory effect, minimum of relevancy for the rest of the attributes, maximum of relevancy for the rest of the attributes, respectively). Yet, this may not be enough (for instance if the twins have $\mu_{rel} = 1$). So, in very rare cases, we could need other refinements, as in the crisp approach.

$${}^2 a \rightarrow_{G1} b \doteq \max\{c \in [0, 1] \mid \min(c, a) \leq b\}$$

$${}^3 a \rightarrow_{Gn} b \doteq \min(1, \frac{b}{a})$$

$${}^4 a \rightarrow_D b \doteq \max(1 - a, b)$$

$${}^5 a \rightarrow_R b \doteq 1 - a \cdot (1 - b)$$

$${}^6 a \rightarrow_L b \doteq \min(1, 1 - a + b)$$

V CONCLUSION

Our approach for diagnosis is mainly causal. It is in-between the probabilistic approach (as we have here uncertain causality, but with no a priori knowledge) and the fuzzy relational approach (more concerned by modelling intensity degrees). This paper generalizes the former approach given by Cayrac et al. [2] (as proved in [5]) to attributes of any kind and is based on fundamental notions of diagnosis in logic: discarding malfunction functions through out a consistency test and suggesting specific malfunctions with an abductive test.

Future work will concern some other refinements of the consistency and abductive possibilistic tests, using some ideas from [6] in order to consider the malfunctions *most* of the *important* effects of which *could* be present. A special attention will be devoted to further refinements of *REL*.

Also, we focused here on single-malfunction detection. For some concrete diagnosis process, the multiple-malfunction detection needs to be considered. It may be done easily if the effect superposition hypothesis holds (in order to avoid having to model the effects of every combination of malfunctions).

Finally, the optimization of the diagnosis process will be studied, following Mo's results ([9]) for optimizing the order in which attributes are considered.

REFERENCES

- [1] K. P. Adlassnig, W. Scheithauer and G. Kolarz, Fuzzy medical diagnosis in a hospital, *Fuzzy Logic in Knowledge Engineering* (H. Prade, C. V. Negoita, eds.), pp. 275–294, 1986.
- [2] D. Cayrac, D. Dubois and H. Prade, Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application, *IEEE Trans. on Fuzzy Systems*, Vol. 4, pp. 251–269, 1996.
- [3] J. Montmain and S. Gentil, Interprétation qualitative pour le diagnostic en ligne, *Revue Européenne Diagnostic et sûreté de fonctionnement*, Vol. 3, n° 1, pp. 23–45, 1993.
- [4] O. De Mouzon, D. Dubois and H. Prade, Une approche Générale du diagnostic causal, *LFA 99*, pp. 222–229, 1999.
- [5] D. Dubois, M. Grabisch and H. Prade, A general approach to diagnosis in a fuzzy setting, *IFSA 99*, Vol. II, pp. 680–684, 1999.
- [6] D. Dubois, M. Nakata and H. Prade, Find the items which certainly have (most of the) important characteristics to a sufficient degree *IFSA 97*, Vol. II, pp. 243–248, 1997.
- [7] D. Dubois, H. Prade and C. Testemale, Weighted fuzzy pattern matching, *IFSA 88*, Vol. 28, n° 3, 1988.
- [8] J. Kitowski and M. Bargiel, Diagnosis of faulty states in complex physical systems using fuzzy relational equations, *Approximate Reasoning in Intelligent Systems, Decision and Control* (E. Sanchez, L. A. Zadeh, eds.), pp. 175–194, 1987.
- [9] X. Mo, Compilation de bases de connaissances avec prise en compte de l'imprécision et de l'incertitude, *Thesis, Univ. of Toulouse*, 1990.
- [10] J. Montmain and S. Gentil, Interprétation qualitative pour le diagnostic en ligne, *Rev. Europ. Diag. et Sur. de Fonctio.*, 3, n° 1, 23–45, 1993.
- [11] E. Sanchez, Solutions in composite fuzzy relation equations: application to medical diagnosis in Brouwerian logic, *Fuzzy Automata and Decision Processes* (M. M. Gupta et al., eds.), pp. 221–234, 1977.