15th Triennial World Congress, Barcelona, Spain

POSSIBILISTIC CAUSAL DIAGNOSIS: FROM SINGLE TO MULTIPLE FAULTS

S. Boverie* D. Dubois** O. de Mouzon* H. Prade****

** DGEI, INSA, 135 a. de Rangueil, 31077 Toulouse Cx 4, France*
*** IRIT, UPS, 118 r. de Narbonne, 31062 Toulouse Cx 4, France*
**** Siemens VDO Automotive S.A.S., 1 avenue Paul Ourliac, BP 1149, 31036 Toulouse Cedex 1, France*

Abstract: A general approach to diagnosis based on fuzzy pattern matching is proposed, making use of consistency and inclusion-based indices in the setting of possibility theory. It first concentrates on single-fault detection, using crisp sets before fuzzy sets. Then the diagnosis is extended to multiple-fault detection, making use of information about normal values of attributes. *Copyright © 2002 IFAC*

Keywords: Diagnosis, engine dynamometer, uncertainty, fuzzy sets, possibility theory

1. INTRODUCTION

A general approach to diagnosis has been studied for engine dyno test benches at Siemens VDO Automotive S.A.S., Toulouse, France. The paper explains the theoretical basis, which can be used in other diagnosis problems. See (Boverie *et al.*, 2002) for details on the development process of the application. Section 2 sets the basis of this approach for the single fault hypothesis: Section 2.1 states the causal diagnosis problem. The use of fuzziness is discussed in the diagnosis process, going from crisp knowledge (Section 2.2) to fuzzy knowledge (Section 2.3). Section 3 shows the interest of information about normal values (Section 3.1) in order to cope with multiple faults with only little more information and hypotheses. This information is then used in the single fault diagnosis (Section 3.2) before being extended to multiple fault diagnosis (Section 3.3).

2. SINGLE FAULTS

Different fuzzy set-based approaches to diagnosis have been proposed for the past 25 years. Some try to identify abnormal situations: E.g., (Montmain and Gentil, 1993). Others seek for possible causes that could explain the observations, e.g., (Kitowski and Bargiel, 1987). This paper is concerned with the second problem. The use of fuzzy relation equations was first suggested in (Sanchez, 1977). Fuzzy relations between effects and faults can model either i) an intensity degree with which an effect is observed (e.g., a *quite* strong fever, where *quite* refers to some specific membership level for the fuzzy predicate "strong"). , or ii) a certainty degree attached to the observation, or to the presence of an effect, when a specific malfunction occurs. The approach developed in (Cayrac *et al.*, 1996) for binary attributes is based on the latter interpretation for malfunctions. It introduces indices giving to what extent some (possibly uncertain) observations discard (resp. suggest) the presence of malfunctions (whose effects are sometimes uncertain). See (Dubois *et al.*, 1999) for a formal generalization to non-binary attributes.

2.1 *Notations and underlying hypotheses*

Let $\mathcal{M}$ be the set of all (known) possible malfunctions and $\mathcal{A}$ be the set of the n observable attributes: $\{X_1, \cdots, X_n\}$. Let $m \in \mathcal{M}$ and

$i \in \{1, \cdots, n\}$, then π^i_m denotes the possibility distribution, giving the (more or less) plausible values for attribute X_i when malfunction m (alone) is present. Let U_i be the domain of X_i, so $\boldsymbol{\pi^i_m} : U_i \longrightarrow [0, 1]$. $\boldsymbol{\mathcal{K}^i_m}$ will be the fuzzy set corresponding to possibility distribution π^i_m, representing the <u>k</u>nowledge about the effect of <u>m</u> on $X_{\underline{i}}$, also called *effect* (or *symptom*) *of m on* X_i (e.g., if m_1 is present, an effect is that the temperature (X_1) is under $150°C$, Table 1).

The observations may also be imprecise (or uncertain). $\mu_{\mathcal{O}_i} : U_i \longrightarrow [0, 1]$, a possibility distribution, gives the (more or less) plausible values for the observed value of attribute X_i. $\boldsymbol{\mathcal{O}_i}$ denotes the fuzzy set corresponding to possibility distribution $\mu_{\mathcal{O}_i}$, expressing the imprecision (or uncertainty) of the <u>o</u>bservations (coming from sensors): It represents the possible values for attribute $X_{\underline{i}}$.

$\mathcal{K}^i_m$ and $\mathcal{O}_i$ both express imprecision, when they contain more than one element. Yet:

- Imprecision for $\mathcal{O}_i$ can be "controlled": Changing the sensors would give more precise (but may be more expensive) or less precise observations.
- Imprecision on $\mathcal{K}^i_m$, on the contrary, cannot be reduced (or changed) that easily: It depends on the available knowledge about the *System* only.

Note that when 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 m has no (known) effect on X_i, all values will be allowed:

$$\forall u \in U_i, \pi^i_m(u) = 1. \tag{1}$$

On the contrary, $\mathcal{K}_m \subseteq \{1, \cdots, n\}$ will denote the set of attributes on which m has a <u>*k*</u>*nown effect*.

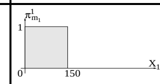
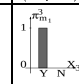
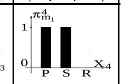
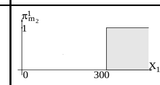
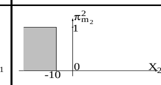
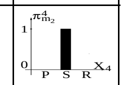
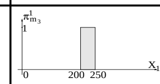
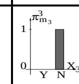
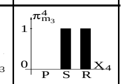
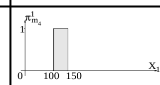
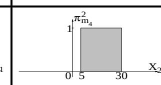
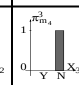
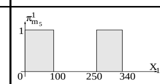
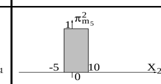
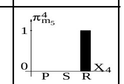
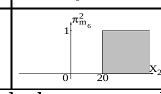
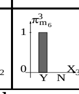
2.2 *Crisp diagnosis*

In this part, only crisp sets are considered in order to better understand how the diagnosis proceeds: $\forall m \in \mathcal{M},\ \forall i \in \{1, \cdots, n\}, \forall u \in U_i, \pi^i_m(u) \in \{0, 1\}$ and $\forall i \in \{1, \cdots, n\}, \forall u \in U_i,\ \mu_{\mathcal{O}_i}(u) \in \{0, 1\}$.

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

This means that for each malfunction, experts were asked very precise bounds, for each attribute, to distinguish values considered as possible from the others (considered as impossible), when the given malfunction occurs. Table 1 gives an example of the type of information experts would give. No graph is provided when experts know nothing about the effects of m_i on X_i, (π^i_m is represented by (1)). On the contrary, if $\exists u \in U_i, \pi^i_m(u) = 0$, then m has an effect on X_i: It restricts the possible values for X_i to $\mathcal{K}^i_m \subsetneq U_i$ ($\bar{\mathcal{K}}^i_m \neq \emptyset$ and $i \in \mathcal{K}_m$ also hold). The more values are forbidden, the more observations may discard m. If $\mathcal{K}^i_m$ contains only one element of U_i, m has a precise effect on X_i (w.r.t. U_i granularity).

Table 1. Crisp Causal Knowledge

$\pi^i_{m_j}$	X_1 ($°C$)	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 when experts know nothing about the effects of m_i on X_i (equivalent to $\forall u \in U_i, \pi^i_m(u) = 1$).

$\mathcal{O}_i = \{u \in U_i \mid \mu_{\mathcal{O}_i}(u) = 1\}$ is the (crisp) set of all possible values for X_i (which are compatible with the observation, o_i, measured on a sensor). This mainly depends on the sensor fiability. The whole paper takes (in the carried examples) 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, there is a conversion from a numerical measurement (o_4) to a symbolic information ($\mathcal{O}_4$). Sensor's reliability gives:

o_4	< 0.95	$[0.95, 0.98[$	$[0.98, 1.03[$	$[1.03, 1.07[$	≥ 1.07
$\mathcal{O}_4$	$\{P\}$	$\{P, S\}$	$\{S\}$	$\{S, R\}$	$\{R\}$

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 m on X_i is present or not.

There are three main cases (and answers):

- $\mathcal{K}^i_m \bigcap \mathcal{O}_i = \emptyset$: The effect is absent.
- $\mathcal{K}^i_m \bigcap \mathcal{O}_i \neq \emptyset$: Effect is present or absent.
- A special case of the latter is $\mathcal{O}_i \subseteq \mathcal{K}^i_m$: The effect <u>is</u> present (it is sure that the effect of m on X_i is present). Yet this does not mean that m is present: For instance, it may happen that in fact m' is present and m absent, if both m and m' have this effect on X_i).

Another special case is the reverse inclusion: $\mathcal{K}^i_m \subsetneq \mathcal{O}_i$ (i.e. $\mathcal{K}^i_m \bigcap \bar{\mathcal{O}}_i = \emptyset$, with $\bar{\mathcal{O}}_i = \{u \in U_i, \mid \mu_{\mathcal{O}_i}(u) = 0\}$, the complement of $\mathcal{O}_i$ in U_i). The test is positive if all the effects of the malfunction *may* have been observed (but not surely). Besides, the $\mathcal{O}_i$ (when o_i is measured) are generally thiner than the $\mathcal{K}^i_m$, so the test is usually negative. Thus this inclusion is not used.

Then, two sets are defined, for malfunctions for which the tests hold for all attributes:

- **CONS** $= \{m \in \mathcal{M} \mid \forall i \in \{1, \cdots, n\}, \mathcal{K}_m^i \bigcap \mathcal{O}_i \neq \emptyset\}$, *all the effects of $m \in CONS$ could be present*: They all are <u>cons</u>istent with the observations.
- **REL** $= \{m \in \mathcal{M} \mid \forall i \in \{1, \cdots, n\}, \mathcal{O}_i \subseteq \mathcal{K}_m^i\}$, *all the effects of $m \in REL$ are present for sure*: They all are <u>rel</u>evant to the observations.

A careful approach to diagnosis would be based on $CONS$ only, which gives all the possibly present malfunctions. E.g., $o_1 = 200°C$, $o_2 = 0Pa$, $o_3 = N$ and $o_4 = 1.05$. Then, $\mathcal{O}_1 = [180, 220]$, $\mathcal{O}_2 = [-3, 3]$, $\mathcal{O}_3 = \{N\}$ 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 really be suspected. Yet, $REL = \emptyset$ (X_1 could be under $200°C$)... Note that the influence of the attribute computation order (in order to optimize the diagnosis time response) is not studied in this paper.

When, $CONS = \emptyset$, no malfunction could be suspected (with respect to the knowledge formalized in the $\mathcal{K}_m^i$). E.g., $o_1 = 274°C$, $o_2 = 10Pa$, $o_3 = Y$ and $o_4 = 0.96$. Then, $\mathcal{O}_1 = [246.6, 301.4]$, $\mathcal{O}_2 = [7, 13]$, $\mathcal{O}_3 = \{Y\}$ 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).

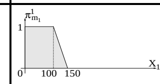
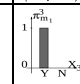
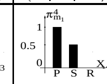
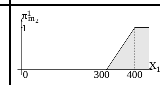
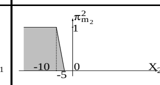
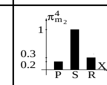
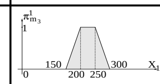
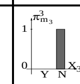
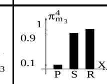
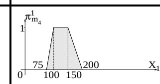
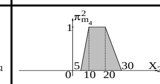
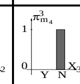
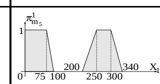
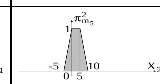
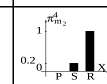
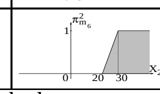
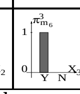
Yet, $CONS$ may select so many malfunctions that it gives no real help to identify a malfunction. In these cases, REL (subset of $CONS$) can be useful as it will list the most likely malfunctions only. E.g., $o_1 = 100°C$, $o_2 = 20Pa$, $o_3 = Y$ and $o_4 = 0.96$, then $\mathcal{O}_1 = [90, 110], \mathcal{O}_2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

REL may discard malfunctions which could be present: It has an abductive behaviour. So $CONS$ is used first. Yet, if $card(CONS) > 1$, REL may help refine the diagnosis and select a more specific malfunction (all the effect of which must have been observed). If an observation is precise (i.e. $card(\mathcal{O}_i) = 1$), $\mathcal{O}_i \subseteq \mathcal{K}_m^i \iff \mathcal{K}_m^i \bigcap \mathcal{O}_i \neq \emptyset$ holds. If all the observations are precise, $REL = CONS$.

2.3 *Introducing Fuzziness*

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 and some others they consider totaly impossible (for all the others, 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

Table 2. Fuzzy Causal Knowledge

$\pi_{m_j}^i$	X_1 $(°C)$	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 when experts know nothing about the effects of m_i on X_i (equivalent to $\forall u \in U_i$, $\pi_m^i(u) = 1$).

then built for representing the given information, is continuous and piecewise linear (see Table 2). For discrete attributes, different levels of possibility may be considered (in Table 2: From 0 to 1 by steps of 0.1 units).

So sets become fuzzy in this (more general) part: $\forall m \in \mathcal{M}$, $\forall i \in \{1, \cdots, n\}$, $\forall u \in U_i$, $\pi_m^i(u) \in [0, 1]$ and $\forall i \in \{1, \cdots, n\}$, $\forall u \in U_i$, $\mu_{\mathcal{O}_i}(u) \in [0, 1]$.

Taking the same example as in the previous section, but defining more precisely the knowledge of the experts together with them, the following knowledge base could be built (Table 2). For instance, the experts will finally say that values for X_1 over $150°C$ are totally impossible when m_1 occurs and that they expect values rather under $100°C$.

The observed values (coming from sensors which only have error tolerances) are still (crisp) sets.

2.3.1. *A Consistency-Based Index*

In this approach a consistency-based index is defined by: $\mu_{cons} : \mathcal{M} \longrightarrow [0, 1]$, 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 $\mathcal{O}_i \bigcap \mathcal{K}_m^i$ is defined by: $u \longmapsto \min(\mu_{\mathcal{O}_i}(u), \pi_m^i(u))$. The elements of highest possibility in this intersection give the *consistency degree between $\mathcal{O}_i$ and $\mathcal{K}_m^i$*:

$$\mu_{cons}^{i,m} = \sup_{u \in U_i} \min(\mu_{\mathcal{O}_i}(u), \pi_m^i(u)). \qquad (2)$$

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):

$$\mu_{\mathbf{cons}}(m) = \min_{i=1}^{n} \mu_{cons}^{i,m}. \qquad (3)$$

The previous crisp solution might look more simple. So one may 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 the corresponding consistency sets be $CONS_{core}$ and $CONS_{support}$ (respectively). Note that $\{m \in \mathcal{M} \mid \mu_{cons}(m) = 1\} = CONS_{core} \subseteq CONS_{support} = \{m \in \mathcal{M} \mid \mu_{cons}(m) > 0\}$.

Yet, $CONS_{core}$ might be empty (although a malfunction should be suspected) and $CONS_{support}$ might select too many malfunctions. E.g., $o_1 = 200°C$, $o_2 = 7.5Pa$, $o_3 = N$ and $o_4 = 1$. Then $\mathcal{O}_1 = [180, 220]$, $\mathcal{O}_2 = [4.5, 10.5]$, $\mathcal{O}_3 = \{N\}$ and $\mathcal{O}_4 = \{S\}$. So $CONS_{core} = \emptyset$ and $CONS_{support} = \{m_3, m_4, m_5\}$. Which malfunction should be preferred? Here, $\mu_{cons}(m_1) = \mu_{cons}(m_2) = \mu_{cons}(m_6) = 0$, $\mu_{cons}(m_3) = 0.9$, $\mu_{cons}(m_4) = 0.4$ and $\mu_{cons}(m_5) = 0.2$. So μ_{cons} prefers m_3. Thus, the use of fuzzy sets allows for a more accurate representation of the knowledge, besides μ_{cons} rank-orders the malfunctions according to their plausibilities. This makes the consistency-based index better in the fuzzy approach than in the crisp one. Moreover, (Dubois *et al.*, 1999) shows how it can be related to the Bayesian approach, under ("natural") conditions.

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: $o_1 = 230°C$, $o_2 = 3Pa$, $o_3 = N$ and $o_4 = 1.1$. Then $\mathcal{O}_1 = [207, 253]$, $\mathcal{O}_2 = [0, 6]$, $\mathcal{O}_3 = \{N\}$ and $\mathcal{O}_4 = \{R\}$. So $\mu_{cons}(m_3) = \mu_{cons}(m_5) = 1$ and the other malfunctions have $\mu_{cons}(m) < 1$.
Thus, as in the crisp case, another index is required in order to refine μ_{cons} and find the most observation-relevant malfunction, among the undiscarded ones, using an abductive approach.

2.3.2. *Abduction-Based Indices* The aim of this section is to single out observation-connected malfunctions. The idea is to use the previously defined (abductive) REL. Yet, the fuzzy aspect of the data requires to define a fuzzy inclusion. This will define an abductive index, as shown below.
In (Dubois *et al.*, 1999), two indices are given for this purpose, representing, in fact, the fuzzy inclusions of $\mathcal{K}_m^i$ in $\mathcal{O}_i$ and of $\mathcal{O}_i$ in $\mathcal{K}_m^i$.

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.
In the following, only fuzzy inclusion of $\mathcal{O}_i$ in $\mathcal{K}_m^i$ 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 $\mathcal{K}_m^i$. The following abductive scheme then holds: $\frac{\mathcal{O}_i,\ m \Rightarrow \mathcal{K}_m^i}{m}$, as the fuzzy inclusion guarantees that if the value is observed in $\mathcal{O}_i$, it is also observed in $\mathcal{K}_m^i$. The extent to which the fuzzy inclusion holds tells how much the abductive scheme holds:

$$\mu_{\mathbf{rel}}^{\mathbf{i,m}} = \inf_{u \in U} \mu_{\mathcal{O}_i}(u) \rightarrow_f \pi_m^i(u). \qquad (4)$$

For n attributes, the weakest fuzzy inclusion tells the extent to which the malfunction is likely:

$$\mu_{\mathbf{rel}}(m) = \min_{i=1}^{n} \mu_{rel}^{i,m}. \qquad (5)$$

An appropriate fuzzy implication ($\rightarrow_f$) had to be chosen. Several of them have been considered, (De Mouzon *et al.*, 2000*a*). Dienes' strong implication ($a \rightarrow_D b \doteq \max(1-a,\ b)$) has been chosen because it is one of the most discriminating for which a parallel may be drawn between (μ_{cons}, μ_{rel}) and their crisp versions:
- $REL \subseteq CONS$: $\mu_{rel} \leq \mu_{cons}$ (if π_m^i is normalised, which is true as 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 $\mathcal{O}_i \bigcap \mathcal{K}_m^i$ are singletons, where the only element u has a possibility degree of $\pi_m^i(u)$).

Going back to the latest example, where μ_{cons} was 1 for both m_3 and m_5, $\mu_{rel}(m_3) = 0.94$ (reached with $X_1 = 253°C$) and $\mu_{rel}(m_5) = 0.14$ (reached with $X_1 = 207°C$). So, m_3 will be suspected first.

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 (Dubois *et al.*, 1988) 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, discrimin or leximin could be used, (De Mouzon *et al.*, 2000*b*). An even more precise knowledge representation can also be defined, using twofold fuzzy sets, (De Mouzon *et al.*, 2000*a*), or weights linked to the different symptoms (e.g., telling how much they necessary follow the presence of a given malfunction).

3. MULTIPLE FAULTS

This section discusses the practical handling of multiple faults, which is often left aside, since it has not the computational simplicity of the "single fault assumption". By multiple faults is meant a combination of faults that are present simultaneously. E.g., someone may have both measles and white tonsillitis. Theoretically speaking, the approach to single-fault diagnosis can detect and identify multiple faults as well, by describing, for each possible combination of faults, its effects on the attributes. Yet, if all sets of multiple faults are possible, one would have to define the associated effects for 2^n faults, if n is the number of single faults. Such a knowledge base would be difficult to obtain in practice and would be quite redundant as in many cases the effects linked to multiple faults are just the "sum" of the effects linked to the different single faults. E.g., measles gives fever and read spots on the body skin, and white tonsillitis gives fever (also) and white spots in the throat. Then, if someone suffers from both measles and white tonsillitis, fever, red spots on the body skin and white spots in the throat are expected.

The following proposes some simple ideas for dealing with multiple faults (at least in an approximate way). The aim is to keep representing the effects of single faults only and to compute the effects of multiple faults (even in a rather imprecise way). First, some information about normal values of attributes is needed (Section 3.1) in order to enable multiple-fault diagnosis. In fact, the single-fault diagnosis can also benefit from this information (Section 3.2). Finally, the approach is extended to multiple-fault diagnosis (Section 3.3).

3.1 *Information about normal values*

The simplest hypothesis, called "*superposition hypothesis*", states that effects of a multiple fault are just the sum of the effects of the single faults involved in the multiple fault, i.e. the effects of all the involved single faults are present. Then, the presence of each multiple fault can be tested with the same indices μ_{cons} and μ_{rel}. The idea is to look for sets of faults with minimal cardinality (in practice one fault sets, first, then two faults sets and so on) whose joint effects are consistent and relevant w.r.t. the observations. This procedure is inspired from the parsimonious covering procedure, first suggested in (Peng and Reggia, 1990). Yet, let m_1 have a known effect on X_1 only and m_2 on X_2 only. Then $m_3 = \{m_1, m_2\}$ would have an effect on X_1 and X_2: $\mathcal{K}^1_{m_3} = \mathcal{K}^1_{m_1}$ and $\mathcal{K}^2_{m_3} = \mathcal{K}^2_{m_2}$. So m_3 would never be selected as the procedure would always prefer m_1 and m_2.
In fact, it is not fully satisfying to represent that m has no (known) effect on attribute X_i by (1). Indeed, either m has *no effect* on X_i and X_i should take normal values when m alone occurs, or m has an *unknown effect* on X_i and then (1) correctly represents this information.

Besides, the diagnosis makes sense only when it explains the observed abnormal values and it should explain *all* observed abnormal values. This way, a multiple fault can be detected and preferred to several possible single-faults.
So, some information about (ab)normal values of attributes is needed. Let $\rho_i : U_i \longrightarrow [0,1]$ be the possibility distribution expressing which values of X_i are known as (more or less) abnormal. For short, it is assumed here that the normal behaviour of X_i is defined as $\eta_i = 1 - \rho_i$.

Now, the former knowledge representation may benefit from this new information. When π^i_m is not given (i.e. $X_i \notin \mathcal{K}_m$), it is because the expert states one of the following cases:
• m (alone) has no effect on X_i, so X_i should keep normal values when m occurs:

$$\pi^i_m = 1 - \rho_i = \eta_i. \tag{6}$$

Let $\mathcal{N}_m \subseteq \{1, \cdots, n\}$ denote the set of attributes on which m has *No effect*.
• m could have an effect on X_i but we do not know to which one, which means that no value of X_i is known as impossible (nor feasible) when m alone occurs: (1). Let $\mathcal{U}_m \subseteq \{1, \cdots, n\}$ denote the set of attributes on which m has an *Unknown effect*.

$\mathcal{K}_m$ still corresponds to the other case: attributes on which m has a *Known effect*. So, $\forall m \in \mathcal{M}, (\mathcal{K}_m, \mathcal{U}_m, \mathcal{N}_m)$ is a 3-partition of $\{1, \cdots, n\}$. $\mathcal{U}_m$ should be as small as possible (as every effect of m on the attributes should be known, even only roughly) and $\mathcal{N}_m$ can be very large. Then, this information on π^i_m is used with the former μ_{cons} and μ_{rel} indices: Equations (3) and (5) are computed with $\min_{i \in \mathcal{K}_m \bigcup \mathcal{N}_m}$ instead of $\min_{i \in \{1, \cdots, n\}}$.

3.2 *Using normal values in single fault diagnosis*

This aspect of diagnosis was missing in Section 2. Yet, the following toy example shows the benefits of this new information for the single-fault diagnosis. Let $\mathcal{A} = \{X_1, X_2\}$ and m_1 and m_2 be two faults s.t. $\mathcal{K}_{m_1} = \{1\}$, $\mathcal{N}_{m_1} = \{2\}$ and $\mathcal{K}_{m_2} = \{1, 2\}$, with $\pi^1_{m_2} = \pi^1_{m_1}$ (so $\mu^{1,m_1}_{cons} = \mu^{1,m_2}_{cons}$ and $\mu^{1,m_1}_{rel} = \mu^{1,m_2}_{rel}$), and $\pi^2_{m_2} \leq \rho_2$ (X_2 has abnormal values, when m_2 occurs), while $\pi^2_{m_1} = 1 - \rho_2$. If $\mu^{1,m_1}_{cons} = \mu^{1,m_2}_{cons} = 1$ and $\mu^{\mathcal{O}_2} \leq 1 - \rho_2$ (X_2 has normal values), then the diagnosis selects m_1 (with μ_{rel}) and discards m (with μ_{cons}), as expected.
Now, suppose that $\mu^{1,m_1}_{cons} = \mu^{1,m_2}_{cons} = 1$ and $\mu^{\mathcal{O}_2} \leq \pi^2_{m_2}$ (X_2 takes abnormal values linked with the presence of m_2). The diagnosis should lead to m_2. Yet, with $\pi^2_{m_1} = 1$ and $\pi^2_{m_1} = 1$, as in the approach of Section 2, m_2 and m_1 are both proposed. Using the information on (ab)normal values ($\pi^2_{m_1} = 1 - \rho_2$) leads to selecting m_2 with μ_{rel} ($\mu_{cons}(m_2) = 1$ and $\mu_{rel}(m_2) \geq 0.5$ as $\mu^{\mathcal{O}_2} \leq \pi^2_{m_2}$) and discarding m_1 with μ_{cons} ($\mu_{cons}(m_1) \leq 0.5$ as $\mu^{\mathcal{O}_2} \leq \pi^2_{m_2} \leq \rho_2$ and $\pi^2_{m_1} = 1 - \rho_2$).

So, information on (ab)normal values can be useful in single-fault diagnosis. Yet, it is a priority for multiple-fault diagnosis, as seen in Section 3.1.

3.3 *Extension to multiple fault diagnosis*

From a crisp point of view, the "*superposition hypothesis*" means that the effects of multiple faults on X_i are the intersection of the effects of the corresponding single faults on X_i: $\forall i \in \{1, \cdots, n\}, \mathcal{K}^i_{m_3} = \mathcal{K}^i_{m_1} \bigcap \mathcal{K}^i_{m_2}$, if $m_3 = \{m_1, m_2\}$. Note that when $m \in m'$ and $i \in \mathcal{N}_m$, then m has no effect on X_i and so it has no incidence on the computation of the effect of m' on X_i. If the above situation holds $\forall m \in m'$ for X_i, then m' has no effect on X_i ($i \in \mathcal{N}_{m'}$): $\pi^i_{m'}$ follows (6). Otherwise:

$$\forall m' \subseteq 2^{\mathcal{M}}, \forall i \in \{1, \cdots, n\}, \forall u \in U_i, \\ \pi^i_{m'}(u) = \min_{m \in \{m'' \in m' | i \in \mathcal{K}_{m''}\}} \pi^i_m(u). \quad (7)$$

Note that $\pi^i_{m'}$ is still a fuzzy set and that the minimum in (7) should be based on $m \in \{m'' \in m' \mid i \in \mathcal{K}_{m''} \bigcup \mathcal{U}_{m''}\}$. But the m'' such that $i \in \mathcal{U}_{m''}$ can be left aside as $\pi^i_{m''} = 1$.

Yet, the "superposition hypothesis" is not always acceptable: Some effects cannot superpose as they are contradictory (they lead to $\pi^i_{m'} = 0$). E.g., disease α gives fever, disease β gives hypothermia. What can be computed for the effect on the body temperature when those two diseases are simultaneously present? Will one effect be stronger than the other? or will they both compensate into a normal body temperature? or will this lead to something different, as a very strong fever? Here, the "superposition hypothesis" cannot hold. In fact, all types of effect combination can be computed, provided that the calculation yields a fuzzy set. Another combination of interest could be the worsening of the effects. For instance, if one disease gives fever and an other too, the presence of both diseases might give a *strong* fever.

It might seem that we are back to the initial problem: For every multiple faults, and each attribute, the type of combination must be defined. This is true when there are no means to "guess" the type of combination. Yet, in most of the cases (in the domain where this multiple-fault diagnosis is applied), the "superposition hypothesis" holds, unless it comes to a contradiction. When a contradiction is reached, a general approach consists in stating that m' has an effect on X_i but which cannot be computed precisely: $\pi^i_{m'} = 1$, as in (1). Of course, when this effect combination computation is not satisfying for multiple fault m' on X_i, $\pi^i_{m'}$ can still be defined in the knowledge base.

4. CONCLUSION

This project has developed a diagnosis method handling uncertainty, based on fuzzy pattern matching. Fuzzy sets do improve human knowledge representation and give a better rank-ordering to the diagnosis process. Both static and dynamic systems may benefit from this method. See (Boverie *et al.*, 2002) for an application on a dynamic system. Future work will implement the extension to multiple-fault diagnosis. The diagnosis of "cascading faults" will also be carried on, (De Mouzon *et al.*, 2001).

REFERENCES

Boverie, S., D. Dubois, X. Guérandel, O. de Mouzon and H. Prade (2002). Possibilistic causal diagnosis: Application to engine dyno test benches. In: *Proc. 15th IFAC World Congress.*

Cayrac, D., D. Dubois and H. Prade (1996). Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application. *IEEE Trans. on Fuzzy Systems* **4**, 251–269.

De Mouzon, O., D. Dubois and H. Prade (2000*a*). Using consistency and abduction based indices in possibilistic causal diagnosis. In: *Proc. 9th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE 2000).* pp. 729–734.

De Mouzon, O., D. Dubois and H. Prade (2001). Twofold fuzzy sets in single and multiple fault diagnosis, using information about normal values. In: *Proc. 10th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE 2001).* In press.

De Mouzon, Olivier, Didier Dubois, Henri Prade and Serge Boverie (2000*b*). Causal diagnosis in engine dyno test benches: A possibilistic treatment. In: *Proc. 15th German-French Institute of Automation and Robotics Conference – Intelligent Control and Diagnosis (IAR – ICD 2000).* pp. 27–32.

Dubois, D., H. Prade and C. Testemale (1988). Weighted fuzzy pattern matching. *Fuzzy Sets and Systems* **28**(3), 313–331.

Dubois, D., M. Grabisch and H. Prade (1999). A general approach to diagnosis in a fuzzy setting. In: *Proc. 8th Int. Fuzzy Syst. Assoc. World Conf. (IFSA 99).* pp. 680–684.

Kitowski, J. and M. Bargiel (1987). Diagnosis of faulty states in complex physical systems using fuzzy relational equations. In: *Approximate Reasoning in Intelligent Systems, Decision and Control.* pp. 175–194.

Montmain, J. and S. Gentil (1993). Interprétation qualitative pour le diagnostic en ligne. *Revue Européenne Diagnostic et Sureté de Fonctionnement* **3**(1), 23–45.

Peng, Y. and J. A. Reggia (1990). Abductive inference models for diagnostic problem-solving. In: *Symbolic Computation – AI.* Spring.-Ver.

Sanchez, E. (1977). Solutions in composite fuzzy relation equations: application to medical diagnosis in Brouwerian logic. In: *Fuzzy Automata and Decision Processes.* pp. 221–234.