An Approach to Evaluation on the Accuracy of the Knowledge Bases Applied to Diagnosis Systems

Isabel Aguiló Pons   Gabriel Fiol Roig

Universitat de les Illes Balears.Departament de Ciències Matemàtiques i Informàtica

Crra. de Valldemosa, Km. 7.5, 07071 Palma de Mallorca, Spain

e-mail: *dmiiap0@ps.uib.es*   *dmigfr0@ps.uib.es*

Abstract
The main contributions of this paper consist in stating a proposal for the evaluation and the improvement of the knowledge bases in order to assure their quality.
This paper addresses the issue of analysing selected qualitative properties of the diagnosis systems in a systematic way and designing the set of rules in an appropriate manner. We investigate a common model for the representation of a base knowledge called Object Attribute Table.
The evaluation of a given knowledge base, is measured using the following steps: first we obtain its complete extensional representation, next we deduce the characteristics that an optimal rule base should exhibit. We consider criteria that measure the accuracy of different knowledge bases describing the concepts.
Keywords: evaluation, quality, criteria, accuracy, optimal diagnosis.

1 Introduction

Guaranteeing the quality of a Knowledge Base is a growing research concern. Complexity of advanced information processing systems makes it more and more difficult to guarantee their accuracy. Finding out adequate diagnosis system of concepts in terms of attributes, it is a very explored problem in the area of Artificial Intelligence. Getting a high quality of the response of the diagnostic process is not an easy task, due essentially to their complexity such as the great amount of rules and variables to be observed, and the existence of redundant and incomplete information. An efficient diagnostic process should be capable of guaranteeing an optimal average between the response time and the quality of the results. To do this an adequate knowledge structure including those information elements of the system needed for a complete diagnosis must be implemented.
Several efforts are being developed related to this subject; such approaches can be classified depending on how they focus on the problem, based either on filling some missing attribute values, or on looking at the probability distribution of values of attributes.
In this paper a common model for representing the knowledge, based on one Object Attribute Table (*OAT)* is used; this representation is very similar to tabular system

(Ligeza, 1998) based on the traditional *Relational Databases*. One of the main factors affecting the quality of the resultant diagnosis system is the characteristic of the nature of knowledge explicited in the *OAT*. In order to assure safe of such systems at a reasonable level de quality of the knowledge base the set of rules must be designed in an appropriate way.
Common kinds of knowledge in a lot of applications are described below:

1.1 Completely Specified Extensional Knowledge

This knowledge is extensional in the sense that all the values about the attributes in describing the concepts have been observed and well defined, that is, there are not missing values. It constitutes in fact the most complete specification of the concepts in terms of the attributes, allowing to find an optimal description of them according to a considered optimality criterion.

In (Aguiló- Fiol, 1995) we considered this type of knowledge where some qualitative features of rule-based diagnosis to evaluate and to improve the quality of rule-based diagnosis system have been studied. It is showed how the success of a diagnosis process depends on the quality of its associate knowledge base and how such a diagnosis process could be improved.
From these works the concept of quality has been applied. Other factors having an influence on the characteristics of the knowledge base have a lot to do with the nature of the own attributes in the *OAT*, so a distinction between binary and multivalued attributes is made in (Aguiló- Fiol, 1996).
Next we consider a new type of knowledge (Aguiló-Fiol, 1997) defined in 1.2, where the analysis of the qualitative aspects about the information in the knowledge base is based on the extensional description of the knowledge base. That is, a description in which the implicit information in the initial knowledge base is now explicited.

1.2 Intensional Knowledge

In this case we consider only essential information about the values of the attributes describing the concepts, then some values of the attributes considered not relevant or simply unimportant values may be obviated by the source of information.

The hardest problem in this theory is the trouble to obtain the implicit knowledge of the knowledge base. What is more, the complexity of the resultant extensional description of the knowledge base, do not allow some procedures to be efficient.
In order to avoid these complexity problems, a method to evaluate the quality of a knowledge base on an approximate analysis of the explicit knowledge in the base is presented in this work in the section 3. Note that the results from this analysis will only constitute an approximation with respect to those obtained by analysing the extensional information of the base. Some approaches are given in (Pawlak, 1984; Quinlan, 1993; Fiol, 1999).

1.3 Incompletely Specified Extensional Knowledge

Finally we consider the possibility of finding some values of the attributes that they are not presented or they are not well specified, that is, corresponding to missing values or unknown values; in (Witold, 1981, Quinlan, 1993; Fiol, 1999; Aguiló 1999) the incomplete information was presented, the knowledge base with incomplete information will be studied in the section 4.

1.4 Objectives

The main contribution of this paper consists in putting forward a proposal to analyse and evaluate some qualitative properties in order to define the accuracy of the knowledge base.
The evaluation of a given knowledge base (*KB*), is measured using the following steps: first we obtain its complete extensional representation, that is, all the information contained in *KB*, which is called the *Accumulated Knowledge* from the *KB* and from the analysis of the Accumulated Knowledge from *KB*, we should be able to deduce the characteristics that an optimal rule base should exhibit. Next, we consider criteria that measure the accuracy of different knowledge bases describing the concepts. The optimality criteria describing the characteristics of the desired solution for the particular problem, that is an optimal subset of attributes to describe the concepts must be found. Finally, the quality of a *KB* is calculated by comparing the characteristics of that *KB* to those of an optimal knowledge base.

2 Knowledge Representation

Knowledge representation constitutes an important issue for practical applications of rule-based systems and research in the area of knowledge engineering. Numerous knowledge representation formalism developed so far allows for efficient representation of complex structures.
We define the notion of enlarged Object Attribute Table in order to represent a rule-based system; the basic idea consists in defining knowledge representation structure based on conceptual model.
Let $D = \{d_1, d_2, \ldots, d_m\}$ be an extensionally defined set of elements and $R=\{r_1, \ldots, r_n\}$ an extensionally defined set of attributes (binary or multivalued), such that for each d_j □ D all the values of these attributes are known. The set of attributes R denotes some properties selected for expressing the domain knowledge of the system (diagnosis system). They are aimed at representation of precondition knowledge for the rules.
This information will be stored in the Object Attribute Table (*OAT*), defined as follows:
The *OAT* is defined as a four-tuple as follows:
OAT = <D, R, V, F>, where
$D = \{d_1, d_2, \ldots, d_m\}$ is a set of elements or rules of the diagnosis system.

$R = \{r_1, r_2, \ldots, r_n\}$ is a set of qualities or attributes, corresponding to the variables of the rules.

$V = \{V_1, V_2, \ldots, V_n\}$ is a family of sets, such that V_i is the set of values of the attribute r_i adopted by the elements of D. In data base literature V_i is called the *domain* of r_i.

$F = \{f_1, f_2, \ldots, f_n\}$ is a set of functions that define extensionally the values that each $d_i \in$ D takes for each attribute $r_j \in R$; that is, f_i: $D \times \{r_i\} \rightarrow V_i$, i=1...n. .

In order to describe the output of the rules, we define the function h: $D \rightarrow C$ assigns to each element of D its corresponding subsets C (or concepts), that is, C is a set of subsets of D and h define extensionally disjoint subsets of D.

The figure 1 represents the Enlarged Object Attribute Table (Enlarged OAT); where $t_i^k \in V_k$, k = 1...n, i= 1...m, is the value of the attribute r_k associated to the element d_i through the function f_k and $C_i \in C$ are the subsets of C associated at the element $d_i \in D$, i = 1...m through the function h. Let us observe that the function f defines extensionally disjoint subsets of D.

Now, taking into account the advantage of the uniform scheme of all the rules in the system, the set of rules can be specified in this table represented in the figure 1, each row contains the complete information about one antecedent-consequent rule, that is, it represents the attributes defining the concepts of the last column (the disturbances) and each column can be seen as a function described in extensional form over the attributes. In fact the Enlarged OAT constitutes the Extensional Information of the concepts.

Let $A=\{r_1, r_2, \ldots, r_m\}$ be a set of attributes used to describe a set C of concepts c_i, i=1...k, Let BR_i be a rule base, described intensionally, obtained by induction from an extensional description, whose information is expressed in the form of antecedent-consequent rules, the attributes of the antecedents belonging to set A, and the concepts of the consequent belonging to set C. Sets A and C are said to be the sets of attributes and concepts respectively associated with BR_i

Given a rule-based diagnosis system, we represent the rules of malfunctions that describe the disturbances of the diagnosis system like rules of the OAT in the form antecedent-consequent rules, that is, $R = \{r_1, r_2, \ldots, r_n\}$ is a set of qualities or attributes, corresponding to the variables of the rules (causes of failures) that describing the disturbances or failures, and $C=\{c_1, c_2, \ldots, c_k\}$ the set of disturbances or malfunctions to be diagnosed. Here is an enlarged OAT.

$$
\begin{array}{c|cccc|c}
 & \multicolumn{4}{c}{\overbrace{\hphantom{r_1 \quad r_2 \quad \cdots \quad r_n}}^{R}} & \\
D & r_1 & r_2 & \cdots & r_n & \\ \hline
d_1 & t_1^1 & t_1^2 & \cdots & t_1^n & C_1 \\
d_2 & t_2^1 & t_2^2 & \cdots & t_2^n & C_2 \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
d_m & t_m^1 & t_m^2 & \cdots & t_m^n & C_m
\end{array}
$$

Any rule in the system is assumed to be of the form:

$$(r_1 = t_1^{1}) \wedge (r_2 = t_1^{2}) \wedge \ldots \wedge (r_n = t_1^{n}) \Rightarrow c_t,$$

where a_i, i=1...n, are identifiers of attributes and t_1^{j}, j=1...n, are the values adopted by these attributes. Each expression in the form $(r_i = t_1^{i})$ is called an attribute-value pair and indicates that the attribute r_i adopts the value t_1^{i}, and c_t is the concept denoting the consequent term of the rule.
Note that for a single row all the qualities defining the values specified for the attributes must hold together; thus the logical connective is conjuction.

3 Knowledge Base with Essential Information

When the *OAT* contains only essential information about the values of the attributes describing the concepts, then some value t_i^{j} of the attributes considered not relevant or simply unimportant values might be obviated by the source of information. In these cases, the quality of the knowledge base to be found depends exclusively on the quality of essential information in the *OAT*.
Note that if an attribute does not appear in a rule we can add this attribute to this rule and we assign it the value underscore ("-"). That is, this attribute is not necessary in the rule in order to define the concepts that appear in the consequent of the rule. This information is considered 'non important' in the description of the concepts.

"Non important" attributes

Let "a" and "b" be two Boolean variables which values can be omitted. In these case the evaluating function will not contain these variables, that is, we will define this function depending on a value "-". Sometimes this situation can be confuse because this value "-" is interpreted as a value of the variable. The approach to the solution of this situation consistS in interpreting a status of the variable but not to interpret its value.

Definition 1. We consider the logic operation @, called "absorption" between two Boolean variables "a" and "b", we denoted by a @ b and we read "a absorbs b". This operation is defined as follows:

$$@: \quad \{0, 1, -\} \times \{0, 1, -\} \text{ ------------> } \{0, 1\}$$

- If it is necessary to know the value of both variables then

$$a @ b = ((\text{not}(a) \text{ and not}(b)) \text{ or } (a \text{ and } b))$$

- If it is not necessary to know the value of some variables then

$$a @ b = (a = \text{"-"})$$

We represent graphically the definition 1: as an absortion function:

a	b	a @ b
0	0	1
0	1	0
1	0	0
1	1	1
-	0	1
-	1	1
0	-	0
1	-	0
-	-	1

Proposition 1. Let KB be a knowledge base and A and C its associated sets of attributes and concepts respectively. Let A'={ a_i, a_j, ..., a_s} be a subset of A. Let e = (v_i, v_j, ..., v_s, c) be a tuple of values of the attributes of A', a_i, a_j, ..., a_s respectively and c its associated concept. Let e' = { v_i', v'_j, ..., v'_s, c'} be another tuple of values of the attributes of A' and c' its associated concept.
The tuple e' is to be absorbed by the tuple e *if and only if*

$$c \neq c' \text{ and } v_k @ v'_k = 1, \ \forall k = 1, j, \ldots s$$

Proposition 2. We consider a pair (e, e') of tuple's of values of attributes of A', associated with knowledge base KB.

The pair (e, e') is said subsumed *if and only if* ($v_k @ v'_k = 1$) or ($v'_k @ v_k = 1$), $\forall$ k=1, s.

Before finding the optimal knowledge base one must determine a suitable subset of attributes Rx, in order to describe the extensional knowledge base without any kind of confusion (contradiction). If is the case, then Rx, is called a basis of attributes.

- Confusion cases.
Let KB be a knowledge base and A and C its associated sets of attributes and concepts respectively . Let A'={ a_i, a_j, ..., a_s} be a subset of A. Let r_1 = (v_i, v_j, ..., v_s, c), r_2 = {v_i', v'_j, ..., v'_s, c' }, $c \neq c'$, where the pair (r_1, r_2) is said subsumed, that is, r_2 absorbs r_1.

The proposed classification is somewhat general, but it covers many detailled confusion cases:.

Case 1.

r_1: $(a_i = vi) \wedge (a_j = v_j) \wedge \ldots \wedge (a_s = v_s) \Rightarrow c$

r_2: $(a_i = v'_i) \wedge (a_j = v'_j) \wedge \ldots \wedge (a_s = v'_s) \Rightarrow c'$

Then, this case is said than (a,b) does not constitute a pair of confusion tuples.

Example:					
	r_1:	1	0	$\Rightarrow$	c
	r_2:	1	-	$\Rightarrow$	c'

Case 2.

r_1: $(a_i{=}v_i) \wedge (a_j{=}v_j) \wedge \ldots \wedge (a_s{=}v_s) \quad |\!\Rightarrow \quad c$

r_2: $(a_i{=}v'_i) \wedge (a_j{=}v'_j) \wedge \ldots \wedge (a_s{=}v'_s) \quad |\!\Rightarrow \quad c'$

where (x $|\!\Rightarrow$ y) means than "x" does not imply necessarily "y"
Then, in this case the pair (r_1, r_2) is said subsumed of tolerant confusion, if the concepts c and c' are described by more rules of the knowledge base KB.

Case 3.

r_1: $(a_i{=}v_i) \wedge (a_j{=}v_j) \wedge \ldots \wedge (a_s{=}v_s) \quad |\!\Rightarrow \quad c$

r_2: $(a_i{=}v'_i) \wedge (a_j{=}v'_j) \wedge \ldots \wedge (a_s{=}v'_s) \quad \Rightarrow \quad c'$

Then, in this case the pair (a,b) is said subsumed of tolerant confusion, if the concept c is described by more rules of the knowledge base KB.

Example : r_1: 1 0 $|\!\Rightarrow$ c

r_2: 1 0 $\Rightarrow$ c'

We can distinguish the concept c' , but we can not distinguish the concept c.

Theorem 1.. Let KB be a knowledge base with essential knowledge and A and C its associated sets of attributes and concepts respectively. Let A' be a subset of A.

The subset A' is a base of attributes if and only if all pair of tolerant confusion subsumed tuples tolerant the confusions.

4 Knowledge Base with Incomplete Information

It is an unfortunate fact of life that data have often missed attribute values. If some values of the attributes are possibly missed or are defined in a vague form, we refer to incompletely specified elements. Let $r_j \in R$ be an attribute of an OAT with domain of discrete values V_j, there are different manners in order to describe the incomplete value of the attribute r_j in the specific situation. An OAT with values incompletely specified is called «an Incomplete OAT»

Definition 2. Let r_j an attribute whose domain of values V_j is discrete Let Vr_j the value that the attribute r_j has with specific situation.
That is, $V_j \supseteq V_{rj}$ and, if $V_{rj} = \{ v_{ji}, v_{jk}, \ldots, v_{jp}\}$, then the value of the attribute r_j is a disjunction of the values of $V_{rj} = v_{ji} \vee v_{jk} \vee \quad \vee v_{jp}$. The attribute r_j is called a discrete attribute. That's mean, $(r_j{=}\ v_{ji},) \vee (r_j{=}v_{jk}) \vee \ldots \vee (r_j{=}\ v_{jp})$

Definition 3. We define the weigth of an incompletely specified attribute, denoted by $Wt(r_j)$, as the number of elements of the subset V_{rj}.
It is important to note that the completely specified attributes have the weigth equal 1. If for some element there is an attribute that has a missing value or unknown value can be assumed with marked, we assign to it the value "*" in the OAT. That is, the OAT is not completely specified.

Example 1 Consider the following knowledge base BR. Let $A=\{a_1, a_2, a_3\}$ be the set of attributes where the values adopted by these attributes are $D_1= \{a,b,c\}$, $D_2=\{0,1,2\}$, $D_3=\{0,1,3\}$ Let $C=\{c_1, c_2, c_3, c_4\}$ be the set of concepts.
The object attribute table (OAT) corresponding to knowledge base is represented as an incompletely OAT.

	a_1	a_2	a_3	C
d_1:	a	0	1	c_1
d_2:	b	1	*	c_2
d_3:	c	*{0,1}	1	c_3
d_4:	a	0	0	c_4
d_5:	*{a,b}	2	1	c_3

The value *{a,b} of the element d_5 represents the value incomplete specified adopted by the a_1 attribute with respect to the concept c_4 of the OAT of the figure 3.
The value * of the element d_2 represents different set of values incomplete specified adopted by the a_3 attribute with respect to the concept c_2 of the OAT of the figure 3, thus *{0,1}, *{0,3}, *{1,3}, *{0,1,3} represent different states about of the OAT.

Proposition 3. Let r_j an attribute whose domain of values V_j is discrete. Let V'_j a value of domain V_j, That is, $V'_j \subset V_j$.

It is said that the value V'_j is a contradictory value *if and only if* the number of elements of the subset of concepts associated to V'_j is bigger than one.

Proposition 4. Let r_j an attribute whose domain of values V_j is discrete. Let V^s_j, and V^t_j two different values of domain V_j, as $V^s_j \subset V_j$, $V^t_j \subset V_j$.

It is said that the values V^s_j, and V^t_j are two-values contradictory *if and only if* $[(V^s_j, \cap V^t_j \neq \varnothing)$ and ((V^s_j, is a contradictory value) or (V^t_j is contradictory value))].

Theorem 2. Let KB be a knowledge base with incomplete information and A its associated sets of attributes. Let A' be a subset of A.

The subset A' is a base of attributes *if and only if* all pair of tuples of values of domain V_j , $[(V^t_j\ V^t_j), (V^t_i\ V^t_i),\ldots, (V^t_k V^t_k)]$ are not contradictory.

5 On the Evaluation of the rule-base

The factors more important on the quality of the resultant diagnostic system are the characteristics of knowledge base make explicit in the OAT. Before finding an optimal intensional description it is very important to determine a suitable subset R_x of the attributes without any kind of confusion or contradictions, this subset R_x is called a basis of attributes. Inducing subset properties from incompletely specified elements suppose the development of suitable induction methods to process the information about the elements, so that the final description of the properties satisfied by the subsets is a adequate as possible according the nature of the problem in hand.

The adequacy of the information in one given knowledge base must be measured. This task has already been formulated in (Aguiló-Fiol, 1995, 1996) whose general algorithm is presented below:

Algorithm Evaluation the quality of rule-based diagnostic system.

1. To establish the knowledge base to be evaluated (KB).
2. We consider the extensional representation associated to the basis of attributes.
3. To define an optimality criterion.
4. To generate an optimal intensional description from the established criterion of the third step (KBO).
5. To compare the (KB) and (KBO) according to the considered criterion.

In the third step, the criterion describing the characteristics that a knowledge base must satisfy to be considered an optimal one is defined. Next a brief description of some optimality criteria is shinthesized:

The minimum cost base criterion. This criterion defines the characteristics of a knowledge base which allows the description of a set of concepts with a minimum cost.- The minimum base criterion. This criterion defines the characteristics of a knowledge base which describe a set of concepts with number of attributes.- The fast base criterion. An optimal knowledge base obtained according this criterion, allow us to classify a set of concepts in fated way.- The minimum time criterion. This criterion is based on the time needed to carry out a diagnosis system of disturbances, this time is the sum of the time to diagnose each branch of the decision tree. An optimal decision tree describe the concepts in with a minimum time.In the fifth step the results obtained by applying the considered criterion to the extensional representation are obtained. The results describe the characteristics that an optimal rules base should exhibit. A computer program UIB-IK[6], developed from inductive techniques, is applied to real time diagnostic system in order to obtain an optimal descriptions of disturbances in the form of a decision tree structure, that is, the program generates an optimal decision tree satisfying a given criterion.

5.1 Accuracy Measures

The problem of measuring the uncertainty of a set of events is not new (Hernàndez-Recasens, 1998; Zadeh, 1968). We intend to study the concept of accuracy from the uncertainty in the observations of the elements. The accuracy of a model is determined through the Shannon entropy which determines the uncertainly of tables incompletely specified.

$$UNC(TOA)= \Sigma_{k=1..m} H_{dk}$$

The Shannon entropy relative to one rule or element is calculated by the equation as the sum of the entropy relative to every attribute:

$$H_{dk}= \Sigma_{j=1..n} H((aj=vi) / c= value)$$

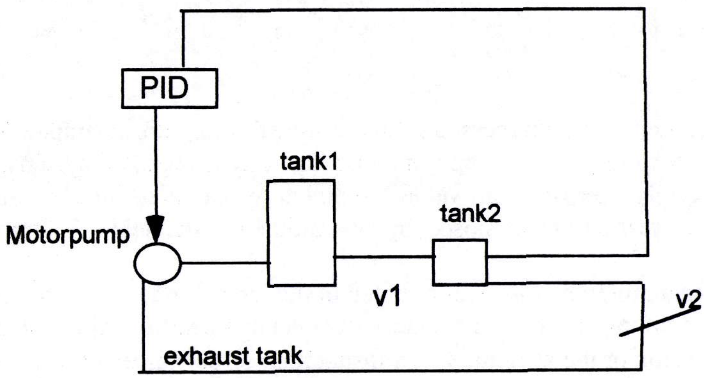
Where P((aj=vi) / (c=value)) is the conditional probability of a certain attribute state to occur, given that the consequent (c=value) has already occurred. The term probability denotes the quotient of the observed frequency of a particular state divided by the highest possible frequency of that state.

$$H((aj=vi) / c=value) =- \Sigma_{i= 1..Wt} P ((aj=vi) / c=value) \log[P ((aj=vi) / c=value)]$$

It is important to note that the completely specified attributes have the weight equal 1, that is, $Wt(r_j)=1$, therefore the Shannon entropy relative to one completely attribute is 0.

6 Experimental Results

In this section we present an example of evaluating of diagnosis system about a Laboratory plant. It is about a system made up of two tanks, interconnected through a control valve V1, which regulates the level of liquid inside tank1 and tank2 as despicted in figure 4. The goal is to control the level in the second tank by pumping the fluid to the first tank while the liquid flows through valve V2. A model of this plant has been obtained and the diagnostic structure has been developed:



- **Fig. 1**: Laboratory Plant

The *OAT* structure of the experiment contains 582 rules, distinguishing only two operation states of the system: the normal and the abnormal operation states. It is only considered a part of the set of rules in order to evaluate the quality of the knowledge base. We are going to obtain the most accuracy and fastest diagnostic system in order to improve the quality of the system. Using the extensional representation OAT of the table 1, the rows represent observations of the expert person about the behaviour of the system, where the failures o faults are the concepts of the OAT, the variables of the

rules that describing the failures are the attributes. Note that in this table there are some rules with attributes that have unknown
values. The set of variables are a_1 that represents the level difference (level tank2-level tank1), a_2 the error signal, a_3 the state of valve V1 (open=0 or closed=1), a_4 the state of valve V2 (open=0 or closed=1), a_5 represents of state of pump (good o bad), and finally a_6, it is the control (PID= 0 or 1), the table 1 represents the incompletely OAT.

Table 1

D	a1	a2	a3	a4	a5	a6	C
d_1:	0	a	0	*{0,1}	go	0	no
d_2:	1	b	1	0	go	0	abno
d_3:	2	c	0	1	bad	0	abno
d_4:	1	a	1	0	go	1	no
d_5:	0	*{a,b}	0	1	bad	1	abno
d_6:	1	a	1	1	bad	1	abno
d_7:	0	b	1	1	go	0	no
d_8:	0	a	0	1	bad	1	no
d_9:	1	a	1	1	bad	1	abno
d_{10}:	0	b	1	0	go	0	no

- The first step of the process consists in transforming an incomplete OAT to a complete OAT. A complete OAT represented in this way is obtained by substituting the values incompletely specified to values completely specified in the substitutions according to the restrictions imposed by the problem. The table 2 represents this complete OAT

- Let the incompletely OAT represented in the table 1, and let d_i an element with incomplete information, let r_j an attribute incompletely specified whose domain is $V_j \supseteq V_{rj}$, then the value of the attribute r_j is a disjunction of the values of $V_{rj} = v_{ji} \vee v_{jk} \vee v_{jp}$.

To simplify the discussion, the updating time for variables identifying inner nodes of the decision tree has been considered as the time unit. Since the UIB-IK tool generates an optimal decision tree satisfying a given criterion, no decision tree faster may be found.

Inner nodes in the tree correspond to attributes or variables describing in the OAT, and leaf nodes represent the diagnosis normal or abnormal.

Next, we calculate the uncertainty associated to complete OAT of the table 2, as: $\mathrm{UNC(OAT)} = \Sigma_{k=1..m} H_{dk}$

H_{d1}=- 1/2 log 1/2-1/2 log 1/2= log 2

$H_{d2}= H_{d3}= H_{d4}= 0$

H_{d5}=-3/5log 3/5-2/5 log 2/5

H_{d6}= H_{d7}= H_{d8}= H_{d9}= H_{d10}= 0

UNC(OAT) = $\Sigma_{k=1..m} H_{dk}$= log 2+0-3/5log 3/5-2/5 log 2/5+0 =0.2749
The table 2 represents the complete specified OAT associated to the Laboratory Plant of the figure 1.

Table 2

D	a1	a2	a3	a4	a5	a6	C
d_1':	0	a	0	0	go	0	no
d_1'':	0	a	0	1	go	0	no
d_2:	1	b	1	0	go	0	abno
d_3:	2	c	0	1	bad	0	abno
d_4:	1	a	1	0	go	1	no
d_5':	0	a	0	1	bad	1	abno
d''$_5$:	0	b	0	1	bad	1	abno
d_6:	1	a	1	1	bad	1	abno
d_7:	0	b	1	1	go	0	no
d_8:	0	a	0	1	bad	0	no
d_9:	1	a	1	1	bad	1	abno
d_{10}:	0	b	1	0	go	0	no

The computer program developed from inductive techniques is applied to Diagnosis System about a laboratory plant in order to find the fastest diagnostic. The figure 2 represents the optimal tree, where the number of questions to classify all the elements is 24.

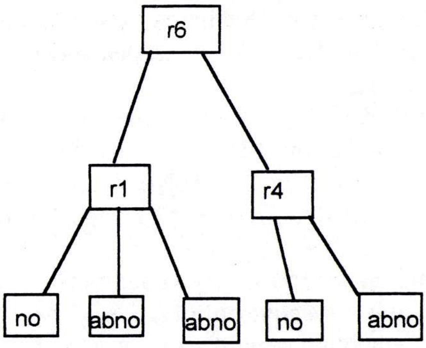


Fig. 2. The optimal tree

The resultant optimal tree generated has an uncertainty degree of a 27%. Once calculated this factor of uncertainty, it is used to establish the accuracy value in order to generate the resultant most accurate tree and the fasted tree.

7 Conclusions

The concept of *quality* of a rules base with *incomplete* information and *essential* information has been introduced and is considered as an measurable concept. It is based on two essential concepts: The *Object Attribute Table*, and *Optimality Criteria*.
An adequate information structure describing the disturbances has been generated. since every analysis of the Knowledge Base must remark in some way the adequate way of describing concepts inside this knowledge base.
The Optimality Criteria describing the characteristics of the desired solution for the particular problem, that is, an optimal subset of attributes to describe the concepts must be found.
A computer tool named UIB-IK, conceived from inductive techniques, has been developed and applied to Diagnosis Systems in particular a Laboratory plant with successful results.
Finally we have studied the concept of accuracy from the uncertainty in the observations of the elements.
Some interesting open points can be stated, particularly those which have something to do with the restrictive conditions demanding identical domains from the rule bases to be evaluated. For instance, extend the results to continous domains and generate fuzzy decision tree.

References

[1] Aguiló, I. and G. Fiol (1995); Qualitative Aspects about System Diagnosis; Advances in System Studies, Vol 2, (pp. 64-68).Baden-Baden. Editor G. Lasker. IIAS – 31-95. ISBN 0921836317

[2] Aguiló, I. and G. Fiol (1996); Evaluation of the Quality of a Rule Base. Proc. of Cybernetics and Systems'96, Vol 2 (pp. 1217-1221). Vienna

[3] Aguiló, I. and G. Fiol (1997); On Incomplete Information in inductive Systems. Invited paper at London. Kmowledge Transfer Vol.II, pp. 38-44. Edited by A. Bedford ISBN 1 900427 15 X. UK.

[4] Aguiló Pons I, Fiol Roig G. (1999) An Approach to the Evaluation of a Diagnostic System with Incomplete and Essential Information. 1999. 7th IEEE international Conference on Emerging Technologies and Factory Automation, Proceedings ETFA'99, pp. 663-670. Spain

[5] Ferrer, M. (1993) UIB-IK, an inductive Acquisition of knowledge System. Master Science Thesis. Computer Science Deparment, Uiniversity of the Balearic Islands. Spain 1993.

[6] Fiol, G. (1999); Inductive Systems with Essential information.Internal Report Balearic Islands University.

[7] Hernàndez, E ; Recasens J.(1998) Observational entropy: entropy in the context of indistinguishability operators.Actas del VIII Congreso Español sobre Tecnologías y lógica Fuzzy. Pamplona 1998.
[8] Ligeza A.(1998). Towards logical analysis of tabular rule-based systems. Proceedings of the Ninth international Worksop, Vienna IEEE, Computer Society. 1998
[9] Pawlak, Z (1984)..On superfluous Attributes in Knowledge representation Systems. Bulletin of the polish Academy of Sciences, 32(3-4).1984.
[10] Quinlan R. (1993); C4.5: Programs for machine Learning. Morgan Kaufman publishers, San Mateo, California, 1993.
[11] Quinlan R. (1993); Unknow Attribute Values, in C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, California ,pp. 27-34.
[12] Witold J.R.(1981); On Databases with Incomplete Information. Journal of the Association for Computing Machinery. Vol. 28. No.I January 19981. pp 41-70.
[13] Zadeh L. (1968) Probability measures of fuzzy events. J. Math. Analysis and Applications. Vol. 23. pp 421-427