Discountings of a Belief Function using a Confusion Matrix

Zied Elouedi
Institut Supérieur de Gestion de Tunis,
LARODEC, Tunisia
Email: zied.elouedi@gmx.fr

Eric Lefèvre
Univ. Lille Nord of France,
UArtois EA 3926 LGI2A, France
Email: eric.lefèvre@univ-artois.fr

David Mercier
Univ. Lille Nord of France,
UArtois EA 3926 LGI2A, France
Email: david.mercier@univ-artois.fr

Abstract—In this paper, we present an analysis of different approaches relative to the correction of belief functions based on the results given by a confusion matrix.

Three different mechanisms based on discountings are detailed. These methods have the objective to assess the discounting rates to be assigned to a source of information. These discounting rates allow to correct raw data, based on learnt decisions given by the confusion matrix.

These corrections differ according to the use of classical or contextual or distance using discountings. An illustrative example is presented to emphasize the interest and also to show the differences between these adjustments.

We carry experimentations on real databases to analyze and interpret these adjustment approaches.

Index Terms—Belief function theory , Belief Function Correction, Discountings, Confusion Matrix.

I. INTRODUCTION

The belief function theory [6], [24], [27] has already shown its ability to manage imperfect data and in particular to deal with information fusion [28], [5], [21].

However, in practice, it is necessary to find a model for building belief functions which will represent available information, but it seems that there does not exist any universal solutions.

Generally, this phase of generation of belief functions is related to the given application and also to the nature of the used sources. Hence, in several applications having a direct link with a real problem, information sources (classifiers, sensors, . . .) are sufficiently considered as known so that all focal elements are fixed and their masses are determined from thresholds or of characteristic abacuses of the used source [4], [11], [13], [20], [23], [30], [31].

In pattern recognition, methods of assignment were defined from a vector relative to the object to be identified, through the calculation of a distance [3], [10], [7], [8], or a likelihood [25], [1], [32], [14]. Methods were also developed for elicitation of expert opinions [2].

When an information source only provides raw decisions and when training data are available, it is possible to convert the given decision to a belief function based on a confusion matrix. This latter includes all past decisions regarding objects where truth is known [15], [18], [33].

In this paper, we handle the case where the confusion matrix is available, and also a method allowing the information source

to provide a belief function, this method is not defined through the confusion matrix, i.e. it is given for example by an expert who is not be influenced by the confusion matrix. The idea is to find how data presented in a matrix can correct or adjust information given by the source.

For this purpose, several solutions are proposed. At first, based on the reliability factors or the percent of correct classification issued from the confusion matrix, a discounting [24, page 252] can be applied. A second approach consists in exploiting the reliability of the source for each context by applying a contextual discounting [17]. Finally, a third proposition will take into account error rates relative to each context.

This paper is organized as follows. In Section 2, basic concepts of the belief function theory will be presented. The definition of confusion matrix is given and illustrated in Section 3. Then, Section 4 proposes the three methods for adjusting the belief function with the confusion matrix, these methods will be detailed and illustrated by an example. Section 5 details and analyzes experimental results obtained from real databases of the U.C.I. repository [22]. Finally, Section 6 gives a conclusion and makes a discussion on this work.

II. BELIEF FUNCTION THEORY

The theory of belief functions is considered as a useful theory for representing and managing uncertain knowledge. This theory is introduced by Shafer [24] as a model to represent quantified beliefs.

In the following, we briefly recall some of the basics of the belief function theory as interpreted in the Transferable Belief Model (TBM). Details can be found in [24], [27].

A. Representation of information

Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$, be a finite set of elementary events relative to a given problem, called the frame of discernment. All the events of Ω are assumed to be exhaustive and mutually exclusive, i.e, only these events are allowed to be handled for the given problem. These events belong to the power set of Ω, denoted by 2^Ω.

For a given agent, the impact of a piece of evidence on the different subsets of the frame of discernment Ω is represented by a basic belief assignment (bba), defined as a function m^Ω :

$2^\Omega \to [0, 1]$ such that:

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1 \qquad (1)$$

If there is no ambiguity regarding the frame of discernment, a basic belief assignment m^Ω can denoted more simply by m.

The mass function m represents the state of knowledge of an agent Ag regarding a question Q. Hence, the value $m(A)$, named a basic belief mass (bbm), represents the portion of belief of an agent Ag committed exactly to the proposition A, and nothing more specific, i.e. *"the answer to the question Q is exactly in the subset A of Ω"*.

The mass $m(\Omega)$ represents the degree of ignorance of the agent Ag. Therefore, the bba defined by $m(\Omega) = 1$ represents the total ignorance. This function is named *vacuous bba* and noted m_Ω. The value $m(\emptyset)$ is the degree of conflict [29]. However, *a certain bba* expresses the total certainty. It is defined as follows: $m(A) = 1$ and $m(B) = 0$ for all $B \neq A$ and $B \subseteq \Omega$, where A is a singleton event of Ω.

For every bba, the subsets A in Ω such that $m(A) > 0$ are called focal elements.

Contrary to the bba which expresses only the part of beliefs committed exactly to A, the belief function bel represents the total belief that one commits to A without being also committed to $\overline{A}$. The belief function bel is defined as follows:

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \text{ for } A \subseteq \Omega \qquad (2)$$

On the other hand, the plausibility function pl quantifies the maximum amount of belief that could be given to a subset A of Ω. Indeed, it contains those parts of beliefs that do not contradict A. The plausibility function pl is defined as follows:

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \text{ for } A \subseteq \Omega \qquad (3)$$

The functions bel, pl et m are in one-to-one correspondence, they represent the same information under different forms.

B. Handling information

1) Combination: Let $m1$ and $m2$ be two basic belief assignments induced from two distinct pieces of evidence, defined on the same frame Ω. These bba's can be combined either conjunctively or disjunctively [29].

If both sources of information are fully reliable, then the induced bba representing the combined evidence is defined as follows:

$$m_1 \textcircled{\scriptsize$\cap$} m_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \qquad (4)$$

On the other hand, if at least one of these sources of information is reliable, we do not know which one, the disjunctive rule is proposed and is defined as follows:

$$m_1 \textcircled{\scriptsize$\cup$} m_2(A) = \sum_{B \cup C = A} m_1(B) m_2(C) \qquad (5)$$

2) Classical discounting of information: A doubt on the reliability of an information source m is sometimes possible. The discounting operation [24, page 252] of m by $\alpha \in [0, 1]$, named *discounting rate*, allows one to take into account this *meta knowledge* on the information m. This correction operation of m is defined by:

$$\begin{cases} {}^\alpha m(A) & = & (1 - \alpha) m(A), \quad \forall A \subset \Omega, \\ {}^\alpha m(\Omega) & = & (1 - \alpha) m(\Omega) + \alpha \;, \end{cases} \qquad (6)$$

or, more simply:

$${}^\alpha m = (1 - \alpha) m + \alpha\, m_\Omega \qquad (7)$$

The coefficient $\beta = (1 - \alpha)$ represents the reliability degree of the information source. If the source is not reliable, this degree β is equal to 0, the discount rate α is equal to 1, and ${}^\alpha m$ is equal to the vacuous bba m_Ω. On the contrary, if the source is reliable, the discounting rate α is null, and m will not be discounted.

Other mechanisms of correction are presented in [12], [16], [34] such that the contextual discounting [17], [19].

3) Contextual discounting of information: Contextual discounting main idea is based on the fact that the reliability of a source of information can be expected to vary according to the truth of the object to be recognized (the context). For example, in target recognition, a sensor in charge of flying targets recognition can be more or less capable of discerning certain types of plane.

A simple method to compute the contextual discounting consists in using its expression through the disjunctive rule of combination (5).

Assume that a source has a degree of reliability β_A knowing that the truth is in A, for each subset A of Ω belonging to a set $\mathcal{A}$ of elements of Ω. Contextual discounting of a piece of information m provided by this source is given by [19]:

$${}^\alpha m = m \,\textcircled{\scriptsize$\cup$}_{A \in \mathcal{A}} A_{\beta_A} \qquad (8)$$

where, for all $A \neq \emptyset$, $v \in [0, 1]$, A_v is another notation [9] for the representation of a bba such that:

$$\begin{array}{rcl} A_v : \quad 2^\Omega & \to & [0, 1] \\ \emptyset & \mapsto & v \\ A & \mapsto & 1 - v \\ B & \mapsto & 0 \;, \; \forall B \in 2^\Omega \setminus \{\emptyset, A\} \end{array} \qquad (9)$$

The bba A_v has two focal elements: $\emptyset$ and A. Besides, $A_{v_1} \textcircled{\scriptsize$\cup$} A_{v_2} = A_{v_1 v_2}$.

Remark 1: As illustrated by (8), within a contextual discounting, each mass $m(A)$ is transferred on $B \supseteq A$ proportionally to the non reliability of the source to recognize elements lying in $A \setminus B$, and its reliability to detect elements in $\Omega \setminus A = \overline{A}$.

Remark 2: If we only know that the source is reliable with a degree β_Ω, the classical discounting is retrieved:

$${}^\alpha m = m \,\textcircled{\scriptsize$\cup$}\, \Omega_{\beta_\Omega} = {}^{\alpha_\Omega} m \qquad (10)$$

Indeed:

$$m \textcircled{c} \Omega_{\beta_\Omega}(A) = m(A)\, \Omega_{\beta_\Omega}(\emptyset) =$$
$$\beta_\Omega m(A) = {}^{\alpha_\Omega} m(A), \quad \forall A \subset \Omega \ , \quad (11)$$

and

$$m \textcircled{c} \Omega_{\beta_\Omega}(\Omega) = m(\Omega)\, \Omega_{\beta_\Omega}(\emptyset) + \Omega_{\beta_\Omega}(\Omega) \sum_{A \subseteq \Omega} m(A)$$
$$= \beta_\Omega m(\Omega) + \alpha_\Omega = {}^{\alpha_\Omega} m(\Omega) \quad (12)$$

C. Pignistic transformation

The problem of decision making in a belief framework as interpreted in the Transferable Belief Model is handled by the pignistic transformation. The TBM is based on a two levels:

- The credal level where beliefs are represented by belief functions.
- The pignistic level where in order to make decisions, belief functions are transformed into probability functions called the pignistic probabilities and denoted BetP [26].

The link between these two functions is achieved by the pignistic transformation:

$$BetP(A) = \sum_{B \subseteq \Omega} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)}, \ \forall A \subseteq \Omega \quad (13)$$

III. CONFUSION MATRIX AND RELIABILITY RATE

In this paper, a confusion matrix without a reject decision is considered.

A *confusion matrix* $\mathcal{M} = (n_{k\ell})_{k \in \{1,\dots,K\}\, \ell \in \{1,\dots,K\}}$ associated to an information source is a table describing the performances of this source on a testing set (Table I).

Each line k corresponds to a decision in favor to ω_k. Each column ℓ corresponds to the case where the truth is ω_ℓ. The general term $n_{k\ell}$ is equal to the number of objects of the class ω_ℓ that have been classified in the class ω_k by the source.

TABLE I
ILLUSTRATION OF A CONFUSION MATRIX.

Decision \ Truth	ω_1	$\dots$	ω_K
ω_1	n_{11}	$\dots$	n_{1K}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
ω_K	n_{K1}	$\dots$	n_{KK}

Note that $n_{k\cdot} = \sum_{\ell=1}^{K} n_{k\ell}$ represents the number of objects classified in ω_k, where $k \in \{1,\dots,K\}$, and $n_{\cdot\ell} = \sum_{k=1}^{K} n_{k\ell}$ the number of objects belonging to ω_ℓ, where $\ell \in \{1,\dots,K\}$. The total number of objects classified in the matrix is given by $n = \sum_{k=1}^{K} \sum_{\ell=1}^{K} n_{k\ell}$.

The *reliability rate* or the *percent of correct classification* $\mathcal{T}_f$ of a matrix is then defined by $\mathcal{T}_f = \frac{\sum_{k=1}^{K} n_{kk}}{n}$.

IV. ADJUSTMENT OF A BELIEF FUNCTION BY A CONFUSION MATRIX

A. Adjustment by classical discounting

A first approach consists in using the reliability rate $\mathcal{T}_f$ of an information source in order to discount the information m provided by this source:

$$^{\alpha} m \ = \ m \textcircled{u} \Omega_{\mathcal{T}_f} \quad (14)$$

B. Adjustments by contextual discounting

1) Use of the percent of correct classification for each context: A second approach consists in exploiting the reliability of the source for each context ω_ℓ, $\ell \in \{1,\dots,K\}$, i.e. such that the truth is ω_ℓ. To find the reliability rate where the truth is ω_ℓ, we look at the column ℓ of the matrix (Table I). This rate, noted $\mathcal{T}_f[\omega_\ell]$, can be then defined as the percent of correct classification such that the truth is ω_ℓ. Formally:

$$\mathcal{T}_f[\omega_\ell] = \frac{n_{\ell\ell}}{n_{\cdot\ell}} \quad (15)$$

Based on these, information provided by this source can be defined in the following manner:

$$^{\alpha} m \ = \ m \textcircled{u}_{\ell=1}^{K} \{\omega_\ell\}_{\mathcal{T}_f[\omega_\ell]} \quad (16)$$

2) Use of a distance: A third approach consists in finding the reliability rate of a source by taking into account for each context ω_ℓ, $\ell \in \{1,\dots,K\}$ of all the decisions given by the source. To this end, a distance [11], based on the Euclidean one, is proposed to compare for each context ω_ℓ, the different decisions ω_k presented by the source of information. The reliability rate $\mathcal{T}_{f_d}[\omega_\ell]$ is then defined as follows:

$$\mathcal{T}_{f_d}[\omega_\ell] = 1 - \sqrt{\frac{1}{2} \sum_{k=1}^{K} \left(\frac{n_{k\ell}}{n_{\cdot\ell}} - \delta_{k,\ell}\right)^2} \quad (17)$$

where $\delta_{k,\ell} = 1$ if $k = \ell$, et 0 otherwise.

The information provided by the source is then discounted by using the equation (16) with reliability rates $\mathcal{T}_{f_d}$.

Remark 3: For each truth ω_ℓ, if only one number of decisions $n_{k\ell,\, k \neq \ell}$ is different from zero, n_{kk} can be or not equal to zero (i.e. $n_{k'\ell} \neq 0$ if and only if $k' = \ell$ or $k' = k$), then the equations (15) et (17) are equivalent.

In fact, in this case $n_{k\ell} + n_{\ell\ell} = n_{\cdot\ell}$, thus:

$$\begin{aligned}
\mathcal{T}_{f_d}[\omega_\ell] &= 1 - \sqrt{\frac{1}{2}\left(\left(\frac{n_{k\ell}}{n_{\cdot\ell}}\right)^2 + \left(\frac{n_{\ell\ell}}{n_{\cdot\ell}} - 1\right)^2\right)} \\
&= 1 - \sqrt{\frac{1}{2}\left(\left(\frac{n_{k\ell}}{n_{\cdot\ell}}\right)^2 + \left(-\frac{n_{k\ell}}{n_{\cdot\ell}}\right)^2\right)} \\
&= 1 - \frac{n_{k\ell}}{n_{\cdot\ell}} = \mathcal{T}_f[\omega_\ell]
\end{aligned} \quad (18)$$

Example 1: Let us consider a source of information S able to recognize 4 types of objects a, b, c et d, where the confusion matrix is given by Table II.

By reading this matrix, the source well recognizes objects of type a (when the object is of type a, the source decides a, which does not mean that when the source decides a, the object should be of type a).

On the contrary, when the object is of type b, the source recognizes it only 6 times, likewise when the object is of

type c. Eventually, the source recognizes 9 times (over 10) the objects of type d.

Decision \ Truth	a	b	c	d
a	8	1		
b		6		
c		2	6	1
d		1	4	9

Based on this example, we have $n = 38$, $\mathcal{T}_f = \frac{8+6+6+9}{38} = .76$, and:

$$\mathcal{T}_f[a] = \frac{8}{8} = 1 \qquad \mathcal{T}_{f_d}[a] = 1 - \sqrt{\frac{1}{2} \, 0} = 1$$

$$\mathcal{T}_f[b] = \frac{6}{10} = .6 \qquad \mathcal{T}_{f_d}[b] = 1 - \sqrt{\frac{1}{2} \, .22} = .67$$

$$\mathcal{T}_f[c] = \frac{6}{10} = .6 \qquad \mathcal{T}_{f_d}[c] = 1 - \sqrt{\frac{1}{2} \, .32} = .6$$

$$\mathcal{T}_f[d] = \frac{9}{10} = .9 \qquad \mathcal{T}_{f_d}[d] = 1 - \sqrt{\frac{1}{2} \, .02} = .9$$

With the rate $\mathcal{T}_f$, only the percent of correct classifications is taken into account, hence the source will have the same degree of reliability .6 for both of types of objects b and c ($\mathcal{T}_f[b] = \mathcal{T}_f[c] = .6$).

With the rate $\mathcal{T}_{f_d}$, the distribution of decisions has an effect, so the source is considered more reliable for objects of type b over objects of type c ($\mathcal{T}_{f_d}[b] = .67$ and $\mathcal{T}_{f_d}[c] = .6$).

Intuitively, the dispersion of decisions leads to less ambiguity on the true kind of object.

Assume that the source provides an information m such that $m(\{a,c\}) = .8$ and $m(\Omega) = .2$. Then, we get:

- by classical discounting $^{\alpha}m = m \, \textcircled{\cup} \, \Omega_{.76}$:

$$^{\alpha}m(\{a,c\}) = .61 \; , \; ^{\alpha}m(\Omega) = .39$$

- by contextual discounting (16) with the rates $\mathcal{T}_f$ (15) $^{\alpha}m = m \, \textcircled{\cup} \, \{a\}_1 \, \textcircled{\cup} \, \{b\}_{.6} \, \textcircled{\cup} \, \{c\}_{.6} \, \textcircled{\cup} \, \{d\}_{.9}$:

$$
\begin{aligned}
^{\alpha}m(\{a,c\}) &= .8(.6.9) = .432 \\
^{\alpha}m(\{a,b,c\}) &= .8(.4.9) = .288 \\
^{\alpha}m(\{a,c,d\}) &= .8(.6.1) = .048 \\
^{\alpha}m(\Omega) &= .8(.4.1) + .2 = .232
\end{aligned}
$$

- by contextual discounting (16) with rates $\mathcal{T}_{f_d}$ (17) $^{\alpha}m = m \, \textcircled{\cup} \, \{a\}_1 \, \textcircled{\cup} \, \{b\}_{.67} \, \textcircled{\cup} \, \{c\}_{.6} \, \textcircled{\cup} \, \{d\}_{.9}$:

$$
\begin{aligned}
^{\alpha}m(\{a,c\}) &= .8(.67.9) = .481 \\
^{\alpha}m(\{a,b,c\}) &= .8(.33.9) = .239 \\
^{\alpha}m(\{a,c,d\}) &= .8(.67.1) = .0535 \\
^{\alpha}m(\Omega) &= .8(.33.1) + .2 = .2265
\end{aligned}
$$

Based on this example, the contextual discounting allows to discriminate the hypotheses b et d: after this discounting the plausibility of b becomes greater than the plausibility of d, $pl(\{b\}) = .52$ and $pl(\{d\}) = .28$ with rates $\mathcal{T}_f$, $pl(\{b\}) = .47$ and $pl(\{d\}) = .28$ with rates $\mathcal{T}_{f_d}$.

V. EXPERIMENTATIONS

A. Methodology

Let B be a database composed of N vectors (objects). Results obtained from the different classifiers are found as follows:

1) From the base B, we take one object V which will be the only object in the test base.
2) The remaining objects will be distributed in two bases with equivalent sizes: B_{train} and B_{conf} (This latter won't be used in this case).
3) The different classification methods used in this paper (K-nearest neighbors, naïve Bayes and decision trees) are based on the base B_{train} for ensuring their training phase i.e B_{train} is considered as a training set.
4) We test the object V in order to evaluate the performances of the different used classifiers.
5) Decisions emerged from the different classifiers are then fused by majority vote in order to make a general decision.

These different steps will be repeated as often as there is objects in the base B. Consequently, each object of the base B will be tested one time.

B. Implementation

To implement our approach, the following steps should be carried out:

1) From the base B, we remove an object V which will constitute to him only the test base
2) The remaining objects of the initial base B will be distributed in two bases with equivalent sizes namely B_{train} and B_{conf}.
3) Apply different classification methods having as an input the base B_{train} for ensuring their training phase.
4) Once this training phase was achieved, we calculate the confusion matrix by testing every classier with the base B_{conf}.
5) Compute the reliability rates: $\mathcal{T}_f$, $\mathcal{T}_f[.]$ et $\mathcal{T}_{f_d}[.]$, defined Section IV, of each classifier from confusion matrices obtained from the last step.
6) Classify the object V by means of every classifier, which gives us a certain belief function.
7) Discount this belief function by using 3 different manners according to the obtained rates of the step 5.
8) Fuse the discounted belief masses of the classifiers by applying the conjunctive rule (17).
9) Use the the maximum pignistic probability rule (13) (giving only one class) in order to classify the object V.

In the same way as with the study of original classifiers, the different operations are repeated as often as there is objects in the base B. The first three phases are common in both approaches in order to be able to compare the results.

Besides, for every tested object V, there are 25 random drawings in order to establish B_{train} and B_{conf}. To measure

the performance of the different methods, we calculate the mean percent of correct classification.

C. Databases

For this paper, we will use three of well-known classifiers namely k-nearest neighbours (KNN), Naive Bayes (NB), and decision trees (DT). Parameters of each of these methods are optimized on the base B_{train}. Obviously other classification methods can also be used.

In our experiments, we have performed several tests on real databases obtained from the U.C.I. repository [22]. A brief description of these databases is presented in Table III.

TABLE III
DESCRIPTION OF DATABASES

Database	Ref	#instances	#attributes
Iris	IR	150	4
Ionosphere	IO	351	34
Wine	WI	178	13
Wisconsin Diagnostic Breast Cancer	WDBC	569	32

D. Results and comments

Results of our proposed approaches are summarized in Tables IV and V, where for this latter table, we have used reliability rates defined in Section IV.

TABLE IV
CLASSIFICATION RESULTS USING KNN, NB AND DT. VALUES BETWEEN PARENTHESES ARE THE NUMBER OF INSTANCES PER CLASSES

REF	KNN	NB	DT
IR (50/50/50)	0.9576	0.9400	0.8928
IO (225/126)	0.8488	0.6410	0.8683
WI (59/71/48)	0.6953	0.9611	0.8656
WDBC (357/212)	0.9266	0.9419	0.9312

TABLE V
CLASSIFICATION RESULTS BY DIFFERENT APPROACHES

REF	Majority Vote	$\mathcal{T}_f$	$\mathcal{T}_f[.]$	$\mathcal{T}_{f_d}[.]$
IR (50/50/50)	0.9517	0.9523	0.9453	0.9453
IO (225/126)	0.8299	0.8300	0.8872	0.8872
WI (59/71/48)	0.9355	0.9557	0.9571	0.9577
WDBC (357/212)	0.9545	0.9545	0.9539	0.9539

By reading these results, we notice that since a classification method has a bad percent of correct classification rate (or rather very weak with regard to two other methods), our proposed approaches give better results than the majority vote. This is the case, for example, with *wine* and *ionosphere* databases.

Furthermore, every classifier, taken in an individual way, is optimized. Indeed, the parameters of these classifiers are obtained by optimizing the percent of classification on the percent of correct classification on the base B_{train}. The results provided by our approaches could be possibly more significant if classifiers were not optimized.

Besides, as already mentioned (see Remark 3), when there are only errors between two classes in the confusion matrix, values $\mathcal{T}_f[.]$ and $\mathcal{T}_{f_d}[.]$ are identical. Results are then similar in the case of *Wisconsin Diagnostic Breast Cancer* and *ionosphere* databases where only two classes are in competition.

Finally, results obtained from these two approaches are also identical in the case of the *iris* base because, even if this base consists of three classes, the confusion exists only between two of them (*versicolor* and *virginica*).

Hence, the proposed approaches seem to be more robust (in term of the percent or correct classification) when intervenes a little successful classifier within the fusion framework.

In order to study this phenomenon in a more precise manner, we replace on of the used classifier by a "virtual" classifier for which we control its percent of correct classification. This "virtual" classifier is obtained as follows. Knowing the labels of the testing vectors (objects of the testing base), we will modify a certain number according to the value of the expected percent of correct classification.

The vectors which the label is changed are drawn lots. In the same way, the new labels of these vectors are also chosen in a random way. We consider all the obtained labels as stemming from our "virtual" classifier. So, the built "virtual" classifier will replace the KNN classifier in order to have always 3 classifiers to simplify the use of the majority vote.

To study the behavior of the different approaches, the percent of correct classification of the "virtual" classifier will vary between $[0.5; 0.95]$ when we deal with a base of possessing data of 2 classes (*Wine* and *Wisconsin Diagnostic Breast Cancer* databases) and between $[\frac{1}{3}; 0.95]$ when we handle a base of possessing data of 3 classes (*Iris* and *Ionosphere* databases). Variation step is fixed to 0.5. The obtained results are represented on figures 1, 2, 3 and 4.

We can also notice that the results with the methods of contextual discountings are similar for bases of data possessing of 2 classes (see Fig. 3 and Fig. 4).

For *Wine*, *Ionosphere* and *Iris* bases (see Fig. 1, Fig. 2 and Figure 4), we notice that when the percent of correct classification of our "virtual" classifier is weak. Then, the proposed methods are more successful than the majority vote method.

The difference between these approaches becomes blurred especially due to the fact that the "virtual" classifier is of new successful. This analysis is not true for *WDBC* base. This maybe explained by the high percent of correct classifications of the two other classifiers (NB and DT).

So, these various tests show the efficiency of the approaches proposed during the failure of a classifier.

VI. CONCLUSION

In this paper, three methods were presented to correct an information from a matrix of confusion, these methods are based on discounting where reliability rates are computed from this matrix.

We remind that the method with which the source supplies a belief function was not learnt from the matrix of

confusion which constitutes the second distinct information. In the opposite case, it would be necessary to use more careful mechanisms, for example inspired by the cautious rule Denøeux [9] in order to avoid to count several times the same information.

To discount the information, we have applied either global reliability rate (classical discounting) or a reliability rate for every context. It would be also possible to use intermediary reliability rates (based on distance).

Experimental results have shown the interest of these approaches especially by comparing them to results induced from three classifiers namely K nearest neighbors, naïve Bayes and decision trees.

As a future work related to the use of a more developed confusion matrix is planned like adding a reject decision in the matrix or subsets of decisions and not only singletons, etc.

REFERENCES

[1] A. Appriou. Multisensor signal processing in the framework of the theory of evidence. *NATO/RTA, SCI Lecture Series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems*, RTO EN-7:(5-1)–(5-31), 1999.

[2] A. Ben Yaghlane, T. Denœux, et K. Mellouli. Elicitation of expert opinions for constructing belief functions. *Proceedings of IPMU'2006*, Paris, France, Juillet 2–7, pages 403–411, 2006.

[3] S. Ben Hariz, Z. Elouedi, et K. Mellouli. Clustering approach using belief function theory, *Proceedings The Twelveth International Conference on Artificial Intelligence: Methodology, Systems, Applications, AIMSA2006*, Varna, Bulgarie, 13-15 Septembre pages 162-171, 2006.

[4] I. Bloch. Some aspects of dempster-shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters*, 17(8):905–919, 1996.

[5] F. Delmotte and G. Gacquer. Detection of defective sources with belief functions *Proceedings of IPMU'08*, Torremolinos (Malaga), pages 337–344, June 22-27, 2008.

[6] A.P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[7] T. Denœux. A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5):804–813, 1995.

[8] T. Denœux. A neural network classifier based on Dempster-Shafer theory. *IEEE transactions on Systems, Man and Cybernetics A*, 30(2):131–150, 2000.

[9] T. Denœux. Conjunctive and Disjunctive Combination of Belief Functions Induced by Non Distinct Bodies of Evidence. *Artificial Intelligence*, 172:234–264, 2008.

[10] Z. Elouedi, K. Mellouli, P. Smets. Belief decision trees: Theoretical foundations. *International Journal of Approximate Journal*, IJAR 28, (2001) 91-124.

[11] Z. Elouedi, K. Mellouli and P. Smets. Assessing sensor reliability for multisensor data fusion with the transferable belief model. *IEEE Transactions on System Man and Cybernatics - Part B*, 34, 782-787, 2004.

[12] S. Fabre, A. Appriou and X. Briottet. Presentation and Description of Two Classification Methods Using Data Fusion Based on Sensor Management. *Information Fusion*, 2:49–71, 2001.

[13] Z. Hammal, A. Caplier and M. Rombaut. A fusion process based on belief theory for classification of facial basic emotions. *Proceedings of FUSION'2005*, Philadelphie, É-U, Juillet 25–29, paper C1–3, 2005.

[14] É. Lefevre, O. Colot, P. Vannoorenberghe, A classification method based on the Dempste-Shafer's theory and information criteria, *Proceedings of FUSION'99*, pp 1179–1184, 1999.

[15] D. Mercier, G. Cron, T. Denœux and M.-H. Masson. Fusion of multi-level decision systems using the Transferable Belief Model. *Proceedings of FUSION'2005*, Philadelphie, É-U, Juillet 25–29, paper C8–2, 2005.

[16] D. Mercier, T. Denœux and M.-H Masson. A Parametrized Family of Belief Functions Correction Mechanisms. *Proceedings of IPMU'2008*, J. V.E. L. Magdalena and M. Ojeda-Aciego (Ed.), Malaga, Espagne, pages 306–313, 2008.

[17] D. Mercier, B. Quost and T. Denœux. Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextuel Discounting. *Information Fusion*, 9(2):246–258, 2008.

[18] D. Mercier, G. Cron, T. Denœux and M.-H. Masson. Decision fusion for postal address recognition using belief functions. *Expert Systems with Applications*, 36(3):5643–5653, 2009.

[19] D. Mercier. Extending the contextual discounting of a belief function thanks to its canonical disjunctive decomposition. *1st Workshop on Belief Functions*, Brest, France, avril, paper 61, 2010.

[20] N. Milisavljević and I. Bloch. Sensor fusion in anti-personnel mine detection using a two-level belief function model. *IEEE Transaction on Systems, Man and Cybernetics C*, 33(2):269–283, 2003.

[21] F. Perisse, D. Mercier, E. Lefevre and D. Roger. Robust Diagnostics of Stator Insulation Based on High Frequency Resonances Measurements, *IEEE Transactions on Dielectrics and Electrical Insulation*, 16(5):1496–1502, 2009.

[22] MP. Murphy, D. W. Aha. Uci repository databases. http://www.ics.uci.edu/mlearn, 1996.

[23] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik. Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion*, 6(4):283–300, 2005.

[24] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

[25] P. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.

[26] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[27] P. Smets. What is Dempster-Shafer's model? *Advances in the Dempster-Shafer theory of evidence*, R. R. Yager, M. Fedrizzi and J. Kacprzyk (Ed.), Wiley, New-York, pages 5–34, 1994.

[28] P. Smets. Data fusion in the Transferable Belief Model *Proceedings of FUSION'2000*, Paris, France, pages 21–33, 2000.

[29] P. Smets. Analyzing the Combination of Conflicting Belief Functions. *Information Fusion*, 8(4):387–412, 2007.

[30] F. Tupin, I. Bloch and H. Maître. A First Step Towards Automatic Interpretation of SAR Images using Evidential Fusion of Several Structure Detectors. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1327–1343, 1999.

[31] A. Veremme, É. Lefèvre, G. Morvan and D. Jolly. Application of the belief function theory to validate multi-agent based simulations. *1st Workshop on Belief Functions*, Brest, France, avril, paper 123, 2010.

[32] P. Walley and S. Moral. Upper probabilities based only on the likelihood function. *Journal of the Royal Statistical Society*, Series B, 61:831–847, 1999.

[33] L. Xu, A. Krzyzak and C.Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435, 1992.

[34] H. Zhu and O. Basir. Extended discounting scheme for evidential reasoning as applied to MS lesion detection. *Proceedings of FUSION'2004*, Per Svensson and Johan Schubert (Ed.), pages 280–287, juin, 2004.
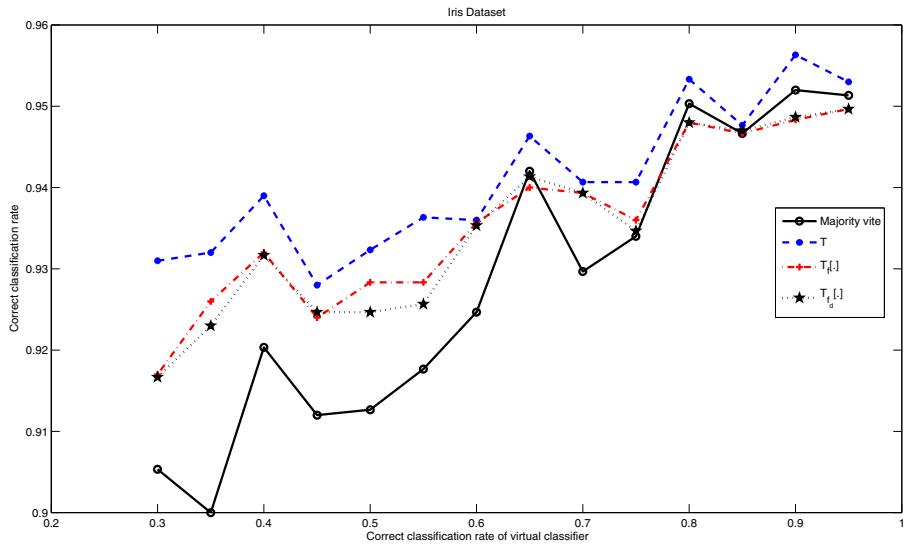
Fig. 1.   Evolution of correct classification for Iris database.
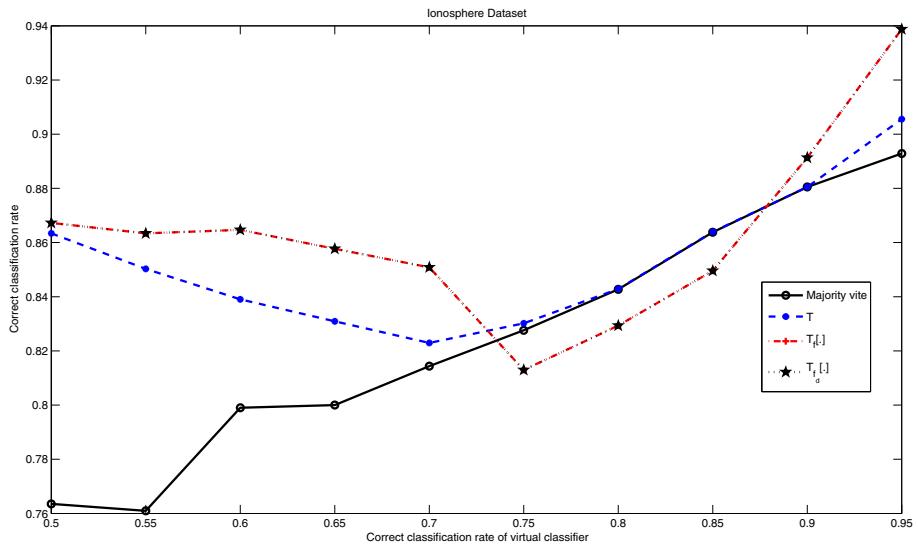

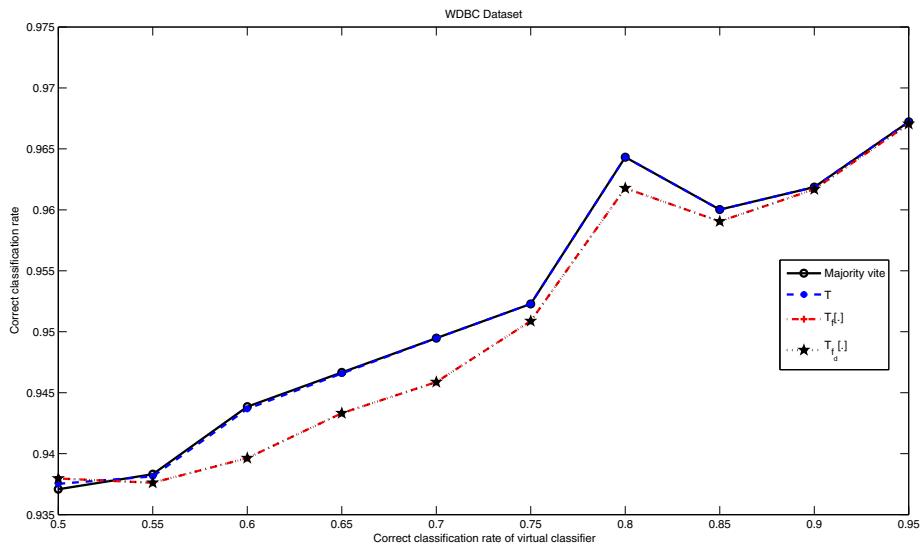Fig. 2.   Evolution of correct classification for Ionosphere database.

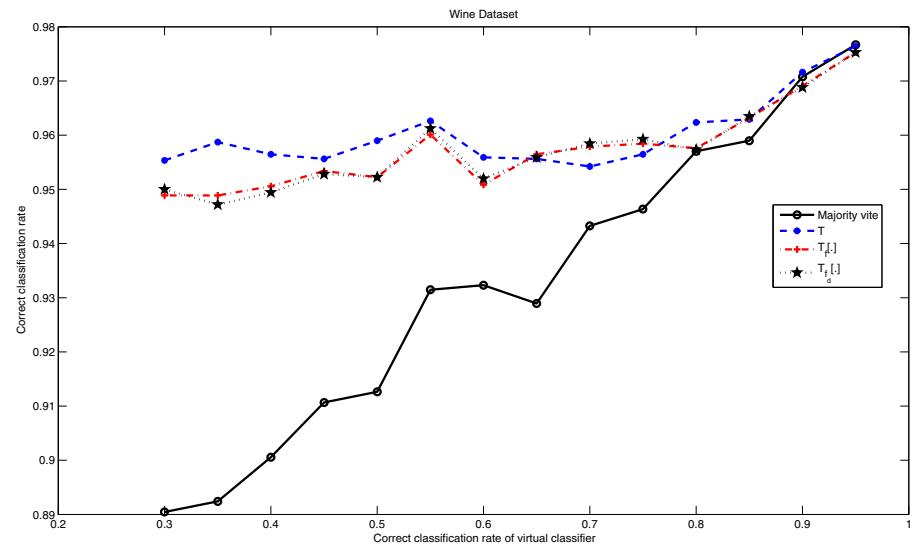Fig. 3.   Evolution of correct classification for WDBC database.

Fig. 4.   Evolution of correct classification Wine database

.