An intelligent quality-based approach to fusing multi-source possibilistic information

Sonda Ammar Bouhamed[1,4], Imene Khanfir Kallel[1,4], Ronald R. Yager[2], Éloi Bossé[3,4], Basel Solaiman[4]

[1]Control and Energy Managment (CEM Lab), University of Sfax,
Sfax Engineering School, BP W, 3038 Sfax, Tunisia
sonda.ammar@gmail.com
imen.khanfir.kallel@gmail.com
[2]Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States
yager@panix.com
[3]Expertises Parafuse & Electrical and Computer Engineering McMaster University,
1280 Main St W, Hamilton, ON L8S 4L8, Canada
ebosse861@gmail.com
[4]Image & Information Processing Department (iTi), IMT-Atlantique,
Technopôle Brest Iroise CS 83818, 29238 Brest Cedex France
basel.solaiman@imt-atlantique.fr

Abstract:

Recently, Yager and Petry were proposing a quality-based methodology to combine data provided by multiple probabilistic sources to improve the quality of information for decision-makers. This paper offers a sort of companion paper that adapts this methodology to possibilistic sources. Possibility theory is particularly well suited to cope with incomplete information from poor-data sources. The methodology and algorithms used for the probabilistic approach are adapted for the possibilistic case. Both approaches are then compared by the means of a numerical example and four experimental benchmark datasets: one, the IRIS data set, being data-poorer than the three other ones (Diabetes dataset, Glass dataset and Liver-disorder dataset). A vector representation is introduced for a possibility distribution as in the probabilistic case and, the Gini's formulation of entropy is being used. However, the Gini's entropy has to be used differently than with the probabilistic case. This has an impact on the selection of subsets. A fusion scheme is designed to select the best-quality subsets according to two information quality factors: quantity of information and source credibility. Results obtained from comparison of both approaches on the four experimental benchmarks confirm the superiority of the possibilistic approach in the presence of information scarcity or incompleteness.

Keywords: Information fusion, quality of information, possibility theory, Gini entropy, source credibility, uncertainty.

1. Introduction

Yager and Petry [1] propose a quality-based methodology to combine data provided by multiple sources in order to improve the quality of information essential to decision-makers in the execution of their tasks. A task can be the estimation of a parameter or, for instance, to perform an inference process about the occurrence of events. Yager and Petry's approach is schematized in Figure 1 using four interrelated functional blocks (a-d):

(a) Modelling of the data and information sources;

(b) Quality-based information criteria;

(c) Ranking of subsets ranking;

(d) Fusion and Subset Selection Process.

The methodology proposed by Yager & Petry is making use of quality-based criteria (block b) in a fusion process (block c) to identify the most valuable set of information (block d) to be used in the execution of a task. For instance, the task is to infer a parametric description for an object, a physical process or an event, given measurements tainted with uncertainty. A measure is being defined to quantify the notion of source credibility that is exploited in the fusion process to provide high-quality fused values for decision-making with reduced uncertainty based on the selection the best credible subset of the sources. Full details concerning Figure 1 blocks (a-d) can be found in [1].

Probability theory is a powerful modelling tool to represent empirical knowledge about random phenomena. The empirical knowledge is generally obtained by sensor observations (hard sensors) and probability theory is an ideal tool to formalize that kind of uncertainty where evidence is based on outcomes with enough independent random experiments. However, in the problem of multi-source information fusion where the information can come not only from hard sensors but also from soft sources such as expert knowledge, contextual knowledge, computing with words [2] and human-source in general [3, 4] , then the information is better modelled either by imprecise probabilities [5], by evidence theory [6], by fuzzy sets [7] and namely, in this paper, by a possibilistic approach [8].

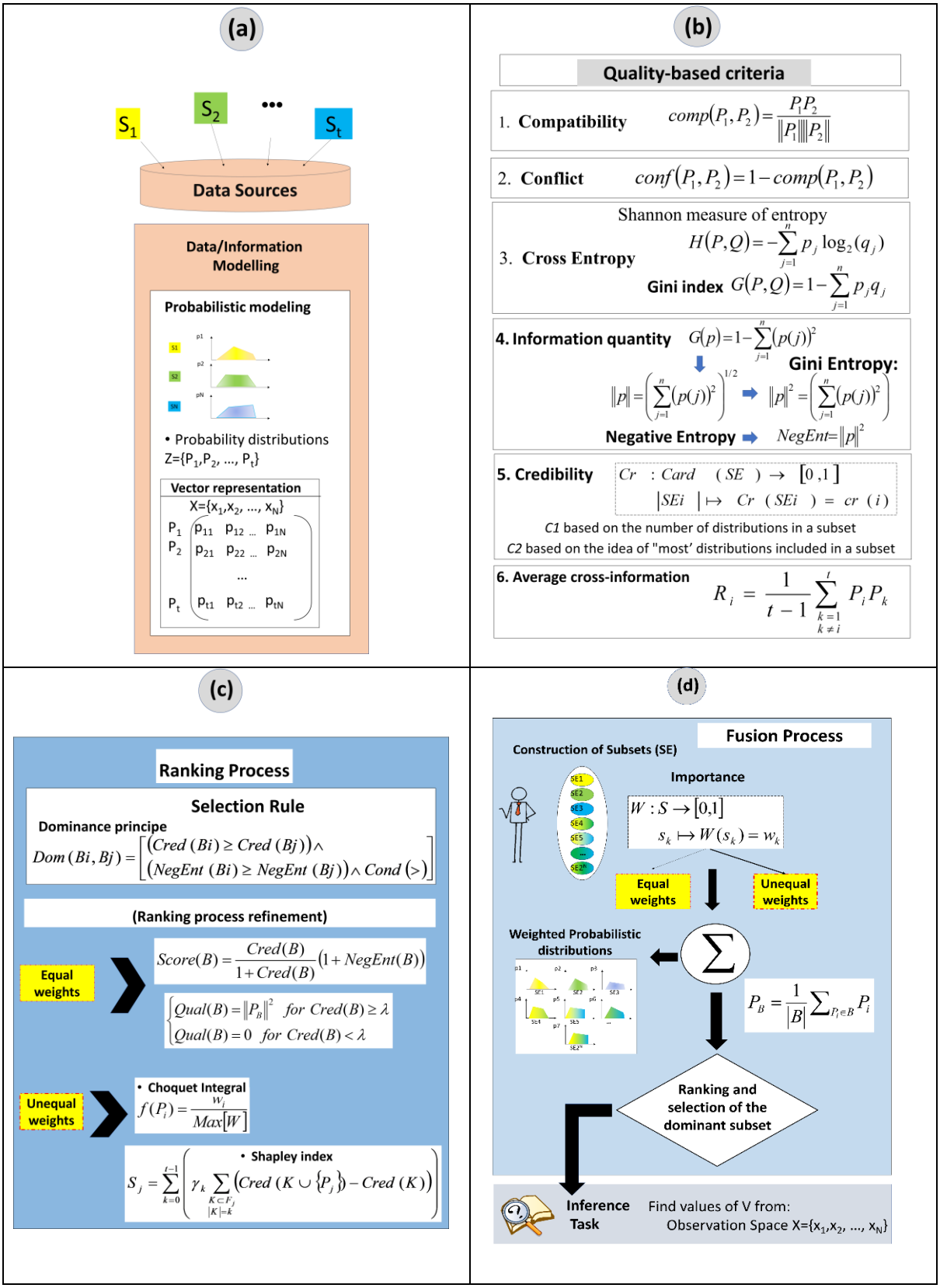


Figure 1. Intelligent quality-based approach of Yager & Petry for probabilistic sources

This paper considers the case of possibilistic sources. It follows the same methodology than in [1] depicted in Figure 1 but the blocks (a-d) have now been translated with the mathematical tools of the possibility theory as shown in Figure 2. In addition, a comparison of both approaches, probabilistic (Figure 1) and possibilistic (Figure 2), is being done through the processing of four experimental benchmark datasets: IRIS, Diabetes, Glass and Liver-disorder.

The paper is organized as follows. Section 2 discusses the related work and scope. Sections 3 to 7 have corresponding sections in [1]. Section 3 is about vector representation of possibility distributions. Section 4 details the use of entropy with possibility distributions. Section 5 is about information in maximally certain and uncertain distribution. Section 6 discusses the uniform fusion of possibility distributions while Section 7 follows on weighted average fusion. Finally, Section 8 presents the results of a comparison of both approaches: probabilitic [1] and possibilitic. For convenience, some important notations with their descriptions are presented in Table 1.

Table 1. List of important notations with their related descriptions.

Notations	**Descriptions**
π	The Possibility distribution
$\pi(j)$	Possibility measure, where $j = 1$ to n
n	The size of the possibility distribution
θ_{12}	The angle between two possibility distributions π_1 and π_2
$comp(\pi_1, \pi_2)$	The compatibility degree between π_1 and π_2
$conf(\pi_1, \pi_2)$	The conflict degree between two possibility distributions π_1 and π_2
$Cred(\pi)$	The credibility measure of a possibility distribution π
SE	The separability measure
P	Probability distribution
$p(j)$	Probability measure, where $j = 1\ to\ n$
$G(\pi)$	The Gini's entropy of a possibility distribution π
$\|\pi\|$	The norm of a possibility distribution π
$\frac{1}{\|\pi\|^2}$	The $NegEnt$ (Negative Entropy) of a possibility distribution π
$\|P\|^2$	The $NegEnt$ (Negative Entropy) of a probability distribution P
w	Weight value
R	A linear aggregation
t	The number of the information sources
Z	A set of possibility distributions
B	A subset of possibility distributions from Z
π_B	The possibility distribution of the subset B
$Dom(B_i, B_j)$	A Boolean variable corresponding to the dominance concept
ND	Non-dominated subsets
m	The number of features

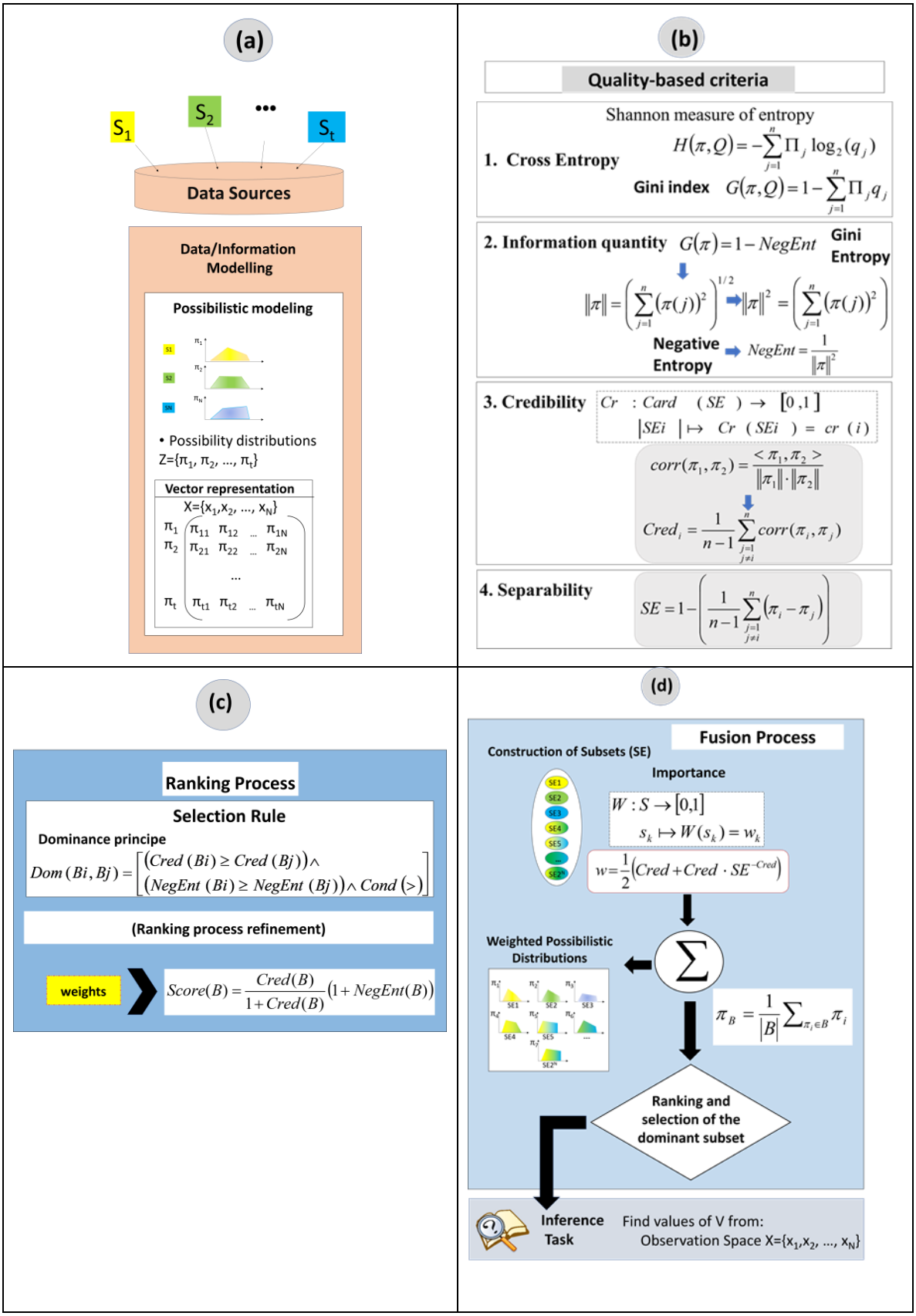


Figure 2. An intelligent quality-based approach for possibilistic sources

2. Scope and Related Work

Several research studies have been proposed in the literature in the field of multi-source data and information fusion systems for low-level fusion (sensors-level) [9-11] and for high-level fusion [4, 12, 13]. This list is far from being exhaustive. Most of these studies are focused on 'fusion systems' and consider much more functions than the 'merging' or the 'combining' rules of a fusion system. Regarding the merging rules, an excellent synthesis of the basic principles for 'combining' pieces of imperfect information, regardless of the representation formalism (sets, logic, partial orders, possibility theory, belief functions or imprecise probabilities), is presented in [14]. They propose to rank the pieces of information to be combined in terms of their relative plausibility as well as to identify impossible values. The relative information content and the mutual consistency of information pieces that affect the performance of the fusion process are also being considered.

When we refer to an "intelligent quality-based" approach, what is meant by "intelligent" is the representation of contextual information items that are being taken into account in the knowledge chain processing from data to information to knowledge to decisions and actions [15]. One of the most complete treatments (gathering-processing-combining-decision-making) along this chain has been presented in [16, 17] for the risk management of natural hazards using beliefs functions and multicriteria decision-making. The book [18] provides a formal foundation and implementation strategies allowing to incorporate information quality into the information fusion processes to various decision support applications for real-life scenarios such as remote sensing, medicine, automated driving, environmental protection, crime analysis, intelligence, defense and security. In [19], the authors describe the whole process of modelling uncertain sensor information. The Dempster–Shafer (DS) theory was chosen to model uncertain sensor information. The authors identify suitable measures from DS theory for determining quality criteria to select subsets on which Dempster's rule of combination is being applied.

Finally, the proposed paper can be considered as a companion paper to [1] in the framework of possibilistic sources. It extends the capacity to treat more types of uncertainty (e.g., fuzziness, vague) and then provides, by the means of a fusion process, a better quality information to decision-makers.

3. Vector representation of possibility distributions

Possibility distributions can also be represented as a n-dimensional vector $\pi = [\pi(1), \pi(2), \dots, \pi(n)]$. Here, π is a normal possibility distribution $\left(\max(\pi) = 1\right)$ on the space $\Omega = \{x_1, x_2, \dots, x_n\}$ where $\pi(j)$ is the possibility of occurrence of x_j. This vector has all its components within the unit interval, $\pi(j) \in [0,1]$. A possibility distribution and its sum can be higher than 1, so that, $\sum_{j=1}^{n} \pi(j) \in [0, M]$ with $M \geq 1$. In

addition, if π_1 and π_2 are two possibility vectors on the space Ω then their dot product, $< \pi_1, \pi_2 >$, a scalar value, is:

$$< \pi_1, \pi_2 > = \sum_{j=1}^{n} \pi_1(j)\pi_2(j) \quad (1)$$

In the case where π_1 and π_2 are normal possibility distributions then $0 \leq (< \pi_1, \pi_2 >) \leq 1$. When π_1 and π_2 are identical, their dot product is:

$$< \pi_1, \pi_2 > = \sum_{j=1}^{n} \left(\pi_1(j)\right)^2 \quad (2)$$

In the following, the dot product will simply be notated as $\pi_i\pi_k$. The norm of a vector is its self dot product, also known as the Euclidean length:

$$\|\pi\| = \sqrt{\pi\pi} = \left(\sum_{j=1}^{n} \left(\pi(j)\right)^2 \right)^{1/2} = (< \pi, \pi >)^{1/2} \quad (3)$$

Using special properties of the possibility distribution vector, $\pi(j) \in [0,1]$ and $1 \leq \sum_{j=1}^{n} \pi(j) \leq n$, it can be shown that:

- **In the case of total ignorance (T.I):** The maximal value of $\|\pi\|$ occurs when all $\pi(j) = 1,\ j = 1\ to\ n$ so $\|\pi\| = \sqrt{n}$ then $\|\pi\|^2 = n$.
- **In the case of complete knowledge (C.K):** $\exists \pi(k) = 1,\ k \in [1, \cdots, n], \forall \pi(j) = 0,\ j \neq k, j = 1\ to\ n$ then $\|\pi\| = 1$ and $\|\pi\|^2 = 1$.

Figure 3 illustrates a possibility vector in the two-dimensional case (n=2). In this figure, each possibility distribution is represented as a point. Non-normal distributions have a vector whose its extremity is "strictly" inside the square. Normal distributions have obligatorily one extremity on a side of the square.

If π_1 and π_2 are two possibility vectors, it is known that the cosine of the angle θ_{12} between them is expressed as in Table 2. This cosine is also the dot product of π_1 and π_2 divided by the vector respective norms. The interpretation of the dot product is similar to what is presented in [1]. As in [1], $\cos\theta_{12}$ can be used as measure of the degree of compatibility, *comp*, between the two possibility distributions; see Table 2. The closer the value of $comp(\pi_1, \pi_2)$ is to one, the more compatible possibility distributions are. Furthermore, $1 - comp(\pi_1, \pi_2)$, denoted $conf(\pi_1, \pi_2)$, is the degree of conflict between two possibility distributions.

- If π_1 and π_2 are orthogonal then $comp(\pi_1, \pi_2) = 0$ and $conf(\pi_1, \pi_2) = 1$.
- If π_1 and π_2 are coincident then $comp(\pi_1, \pi_2) = 1$ and $conf(\pi_1, \pi_2) = 0$.

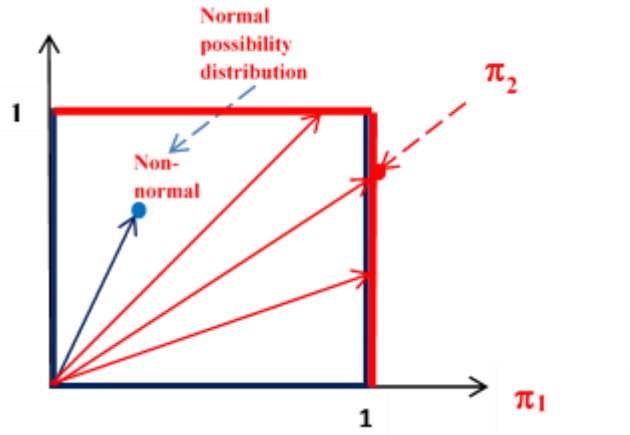


Figure 3. Angle between possibilistic vectors

Table 2. Comparative equations associated with a vector representation under both frameworks

Probabilistic Framework	Possibilistic Framework
$cos(\theta_{12}) = \frac{p_1 p_2}{\|p_1\|\|p_2\|}$	$cos(\theta_{12}) = \frac{\pi_1 \pi_2}{\|\pi_1\|\|\pi_2\|}$
$comp(p_1, p_2) = \frac{p_1 p_2}{\|p_1\|\|p_2\|}$, with $comp(p_1, p_2) \in [0,1]$	$comp(\pi_1, \pi_2) = \frac{\pi_1 \pi_2}{\|\pi_1\|\|\pi_2\|}$, with $comp(\pi_1, \pi_2) \in [0,1]$,
$< P_1, P_2 >= \sum_{j=1}^{n} P_1(j)P_2(j) = \frac{1}{n}\sum_{j=1}^{n} P_2(j) = \frac{1}{n}$	$< \pi_1, \pi_2 >= \sum_{j=1}^{n} \pi_1(j)\pi_2(j)$ $= 1 \sum_{j=1}^{n} \pi_2(j)$, where $\sum_{j=1}^{n} \pi_2(j) \in [0, n]$
$comp(P_1, P_2) = \frac{\frac{1}{n}}{\|P_2\|\left(n^{\frac{1}{2}}\right)} = \frac{\left(\frac{1}{n}\right)^{1/2}}{\|P_2\|}$ $= \frac{1}{\sqrt{n}}\frac{1}{\|P_2\|}$	$comp(\pi_1, \pi_2) = \frac{\sum_{j=1}^{n} \pi_2(j)}{\|\pi_2\|\left(n^{\frac{1}{2}}\right)}$ $= \frac{1}{\|\pi_2\|}\frac{1}{(n^{1/2})}\sum_{j=1}^{n} \pi_2(j)$ $= \frac{1}{\|\pi_2\|}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \pi_2(j)$

We see that the conflict takes its maximal value when π_1 and π_2 are orthogonal and takes its minimum value when π_1 and π_2 are coincident.

- **Complete knowledge (C.K)**: consider π_1 and π_2 being two certain possibility distributions on the same element i like $\exists \pi_1(i) = 1, \pi_2(i) = 1,\ \forall \pi_1(j) = 0, \pi_2(j) = 0\ j \neq i, j = 1\ to\ n$ so $comp(\pi_1, \pi_2) = 1$ and $conf(\pi_1, \pi_2) = 0$.

- **Total ignorance (T.I)**: if one of the distributions, π_1, has $\pi_1(j) = 1, \forall j = 1, \dots, n$.. In this case, $\|\pi_1\| = \sqrt{n}$. Consider now $comp(\pi_1, \pi_2)$ where π_2 is that uniform possibility distribution with $comp(\pi_1, \pi_2) = \frac{\pi_1 \pi_2}{\|\pi_1\| \|\pi_2\|}$. For this case, the comparative equations are given in the last row of Table 2.

There are two special cases of π_2 that worth to be commented.

- **The case of total ignorance (T.I): i**f π_2 is a uniform possibility distribution, all $\pi_2(j) = 1,\ j = 1\ to\ n$. In this case, $\|\pi_2\| = \sqrt{n}$ and $\sum_{j=1}^{n} \pi_2(j) = n$ and $comp(\pi_1, \pi_2) = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}} n = 1$.
- **The case of complete knowledge (C.K):** if π_2 is a certain possibility distribution, $\exists \pi_2(j) = 1,$ for one element, then $\|\pi_2\| = 1$ and $\sum_{j=1}^{n} \pi_2(j) = 1$. In this case, $comp(\pi_1, \pi_2) = \frac{1}{\sqrt{n}}$.

We illustrate three cases of possibility distributions (when n=2) in Figure 4: 1) orthogonality between two distributions, 2) complete knowledge and, 3) the total ignorance distributions. Certain possibility distributions, π_1 and π_2, are completely conflicting and π_3 is an uncertain possibility distribution. Then, in this case, the distributions are: $\pi_1(1) = 1,\ \pi_1(2) = 0,\ \pi_2(1) = 0,\ \pi_2(2) = 1,\ \pi_3(1) = 1,\ \pi_3(2) = 1$.

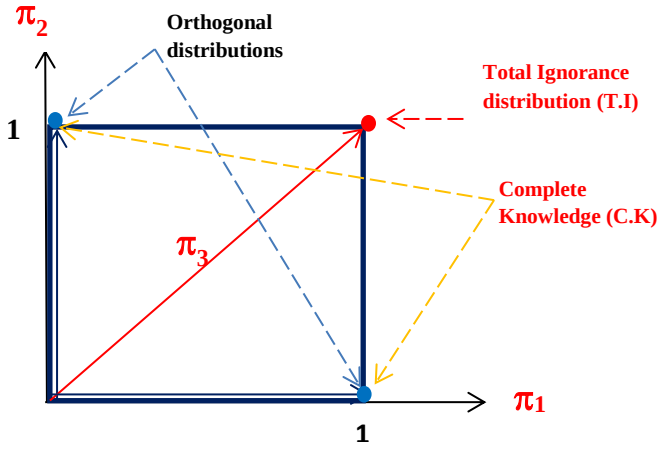


Figure 4. Illustration of orthogonal, complete knowledge and the total ignorance distributions

4. The use of entropy with possibility distributions

Entropy is an important concept for measuring uncertainty associated with a probability distribution. With P, a probability distribution defined on the space $\Omega = \{x_1, x_2, \dots, x_n\}$ with p_j, the probability associated with x_j, then the common measure of entropy is the Shannon entropy defined as:

$$H(P) = -\sum_{j=1}^{n} p(j) \ln(p(j)) \tag{4}$$

According to Klir [20], maintaining a link between a possibility and a probability distribution is important. The uncertainty of a possibility distribution $U(\pi)$ and the entropy of a probability distribution $H(P)$ must be equal, so $U(\pi) = H(P)$.

$$U(\pi) = SP(\pi) + D(\pi) \tag{5}$$

$$SP(\pi) = \sum_{j=2}^{n} \big(\pi(j) - \pi(j+1)\big) \ln j \tag{6}$$

$$D(\pi) = \sum_{j=2}^{n} \big(\pi(j) - \pi(j+1)\big) \ln\left(\frac{j}{\sum_{k=1}^{j} \pi(k)}\right) \tag{7}$$

Finally, $U(\pi) \in [0, \ln(n)]$, and the maximal uncertainty happens when all $\pi(j) = 1$ so $U(\pi) = \ln(n)$. The minimal uncertainty occurs when one $\pi(j) = 1$ and all other $\pi(j) = 0$, hence $U(\pi) = 0$. Larger is $U(\pi)$, larger is the uncertainty. The other side of uncertainty is certainty or information. Smaller $U(\pi)$ is then more information is convoyed by a possibility distribution. This is desirable for decision-making. The maximal entropy occurs when all $p(j) = 1/n$ then $H(P) = \ln(n)$. The minimal entropy occurs when one $p(j) = 1$ and all other $p(j) = 0$ then $H(P) = 0$. For Shannon entropy, we also have $H(P) \in [0, \ln(n)]$. For the same reasons evoked in [1], we use an alternative formulation for entropy called the Gini's entropy defined in [21, 22] for the probabilistic framework and given in Table 3. By analogy, we define the Gini's entropy for a possibility distribution π on Ω by :

$$G(\pi) = 1 - \boldsymbol{NegEnt}, \text{ where } \boldsymbol{NegEnt} = \frac{1}{\|\pi\|^2} \text{ and } \|\pi\|^2 = \sum_{j=1}^{n} \big(\pi(j)\big)^2 \tag{8}$$

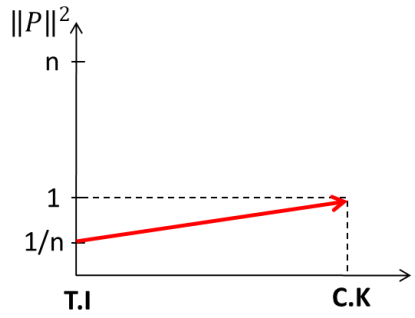
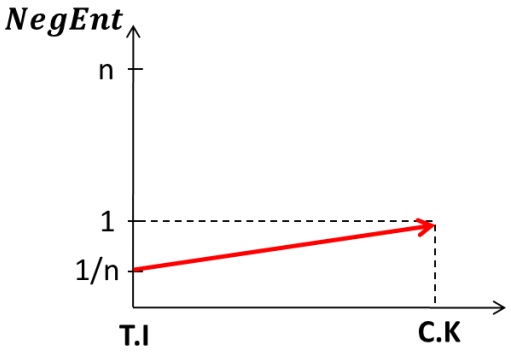
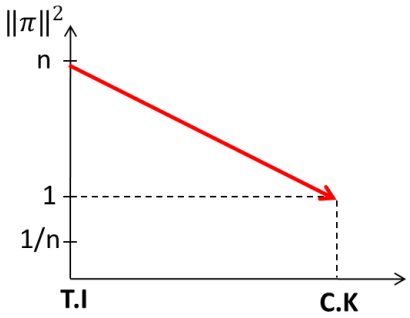
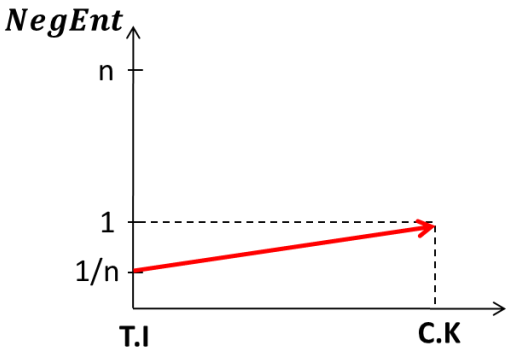
It can be shown that:

- **In the case of total ignorance (T.I):** The Gini's entropy assumes its minimal value $G(\pi) = 1 - \frac{1}{n}$ when all $\pi(j) = 1,\ j = 1, \dots, n.$,
- **In the case of complete knowledge (C.K):** The Gini's entropy assumes its maximal value $G(\pi) = 0$ when only one $\pi(j) = 1$ and all other $\pi(j) = 0$.

The bigger $G(\pi)$ is then the larger uncertainty is conveyed by the possibility distribution. The closer $G(\pi)$ is to zero, the more certainty resides in the knowledge so more information is provided. Thus, to increase certainty or information, we must decrease $G(\pi)$ and, according to equation (8), increase $NegEnt$.

$NegEnt$ is defined differently for probabilistic and possibilistic frameworks, as it is shown in Table 3. However, it has the same variation domain, being $\left[\frac{1}{n}, 1\right]$ and even the same behaviour according to information quantity variations. In both probabilistic and possibilistic frameworks, $NegEnt$ is linked to information quantity.

Table 3. Domains of variation of $\boldsymbol{NegEnt}$ for both frameworks

Probabilistic Framework	Possibilistic Framework
$G(p) = 1 - \boldsymbol{NegEnt}$ where $\boldsymbol{NegEnt} = \|p\|^2$	$G(\pi) = 1 - \boldsymbol{NegEnt}$ where $\boldsymbol{NegEnt} = \frac{1}{\|\pi\|^2}$
Complete Knowledge (C.K): $\boldsymbol{NegEnt} = 1$ Total Ignorance (T.I): $\boldsymbol{NegEnt} = \frac{1}{n}$ $\frac{1}{n} \leq \boldsymbol{NegEnt} \leq 1$	Complete Knowledge (C.K): $\boldsymbol{NegEnt} = 1$ Total ignorance (T.I): $\boldsymbol{NegEnt} = \frac{1}{n}$ $\frac{1}{n} \leq \boldsymbol{NegEnt} \leq 1$
 	 

Let us now examine the variation of π^2 with respect to the quantity of information contained in the possibility distribution. Assume π_1 and π_2 are two possibility vectors on the space Ω and that the relations between them are given as in Table 4 below. In fact, if you replace respectively p_1 by π_1 and p_2 by π_2 as shown in Table 4, the analysis provided in Section 3 of Yager & Petry [1] stands as well for a possibilistic framework.

Table 4. Comparative equations associated with the use of entropy under both frameworks

Probabilistic Framework	Possibilistic Framework
$G(p) = 1 - \sum_{j=1}^{n} (p(j))^2$	$G(\pi) = 1 - \frac{1}{\sum_{j=1}^{n} (\pi(j))^2}$
$p_2(1) = p_1(1) - \alpha$ $p_2(2) = p_1(2) + \alpha$ $p_2(j) = p_1(j)\ for\ j = 3\ to\ n$ For $\alpha \geq 0$. We have $\sum_{j=1}^{n} \big(p_2(j)\big)^2 =$ $(p_1(1) - \alpha)^2 + (p_1(2) + \alpha)^2 + \sum_{j=3}^{n} \big(p_1(j)\big)^2$ Since $(p_1(1) - \alpha)^2 = \big(p_1(1)\big)^2 - 2\alpha p_1(1) + \alpha^2$ and $(p_1(2) - \alpha)^2 = \big(p_1(2)\big)^2 - 2\alpha p_1(2) + \alpha^2$ then we get $\sum_{j=1}^{n} \big(p_2(j)\big)^2 - \sum_{j=1}^{n} \big(p_1(j)\big)^2 =$ $2\alpha \big(p_1(2) - p_1(1)\big)^2 + 2\alpha^2$ $= 2\alpha \Big(\big(p_1(2) - p_1(1)\big) + \alpha\Big)$	$\pi_2(1) = \pi_1(1) - \alpha$ $\pi_2(2) = \pi_1(2) + \alpha$ $\pi_2(j) = \pi_1(j)\ for\ j = 3\ to\ n$ For $\alpha \geq 0$. We have $\frac{1}{\sum_{j=1}^{n} (\pi(j))^2} =$ $(\pi_1(1) - \alpha)^2 + (\pi_1(2) + \alpha)^2 + \frac{1}{\sum_{j=3}^{n} (\pi(j))^2}$ Since $(\pi_1(1) - \alpha)^2 = \big(\pi_1(1)\big)^2 - 2\alpha\pi_1(1) + \alpha^2$ and $(\pi_1(2) - \alpha)^2 = \big(\pi_1(2)\big)^2 - 2\alpha\pi_1(2) + \alpha^2$ then we get $\frac{1}{\sum_{j=1}^{n} \big(\pi_2(j)\big)^2} - \frac{1}{\sum_{j=1}^{n} \big(\pi_1(j)\big)^2} =$ $2\alpha\big(\pi_1(2) - \pi_1(1)\big)^2 + 2\alpha^2$ $= 2\alpha \Big(\big(\pi_1(2) - \pi_1(1)\big) + \alpha\Big)$

5. Information in maximally certain and uncertain distribution

Two probability distributions on X: $P = [p_1, p_2, \cdots, p_n]$ and $Q = [q_1, q_2, \cdots, q_n]$, are considered in Section 4 of Yager and Petry [1]. The objective is to calculate the information associated with the linear aggregation R, that is: $R = w_1 P + w_2 Q$, R is a probability distribution when $w_1 + w_2 = 1$. For each component r_j of R, we have, $r_j = \big(w_1 p_j + w_2 q_j\big)$. The details are in [1]. For the possibilistic case, the information, associated with π, according to the value of the weights w_1 and w_2, is calculated in order to study the impact of the weights on the distribution π. Three different cases are considered:

1. π_1 and π_2 are two certain possibility distributions,
2. π_1 and π_2 are both maximally uncertain distributions, and
3. π_1 is a distribution *completely certain* and π_2 is a distribution *completely uncertain*.

To perform the information associated with π, we calculate:

$$\|\pi\|^2 = \sum_{j=1}^{n} \big(w_1\pi_1(j) + w_2\pi_2(j)\big)^2 = \sum_{j=1}^{n} \Big(w_1^2 \big(\pi_1(j)\big)^2 + w_2^2 \big(\pi_2(j)\big)^2 + 2w_1 w_2 \pi_1(j)\pi_2(j)\Big) \qquad (9)$$

$$\|\pi\|^2 = {w_1}^2 \sum_{j=1}^{n} \left(\pi_1(j)\right)^2 + {w_2}^2 \sum_{j=1}^{n} \left(\pi_2(j)\right)^2 + 2w_1 w_2 \sum_{j=1}^{n} \pi_1(j)\pi_2(j) \quad (10)$$

$$\|\pi\|^2 = {w_1}^2 \|\pi_1\|^2 + {w_2}^2 \|\pi_2\|^2 + 2w_1 w_2 < \pi_1, \pi_2 > \quad (11)$$

We do not repeat here all derivations for the possibilistic representation when they are similar to those of the probabilistic framework. We only list in Table 5 where there are differences. We invite the reader to consult Section 4 of Yager and Petry [1] for details.

Table 5. Comparative equations associated with the calculation of $\|\pi\|^2$

Probabilistic Framework	Possibilistic Framework
$P = \sum_{i=1}^{t} w_i p_i$ with $p(j) = \sum_{i=1}^{t} w_i p_i(j)$ If $w_i \in [0,1]$ and $\sum_{i=1}^{t} w_i = 1$, then p is also a probability distribution vector. (Case when $w_1 = w_1 = 1/2$)	$\pi = \sum_{i=1}^{t} w_i \pi_i$ with $\pi(j) = \sum_{i=1}^{t} w_i \pi_i(j)$ If $w_i \in [0,1]$ and $\sum_{i=1}^{t} w_i = 1$, then π is also a possibility distribution vector. (Case when $w_1 = w_1 = 1/2$)
If P_1 and P_2 are two certain probability distributions, then $\|P_1\| = 1$ and $\|P_2\| = 1$ and there are two cases of interest. 1) If P_1 and P_2 are completely compatible, $P_1(j) = P_2(j) = 1$ for some j. Here $cos(P_1, P_2) = 1$ and $\|P\|^2 = \frac{1}{4}(1) + \frac{1}{4}(1) + \frac{1}{2}(1) = 1.$ 2) If P_1 and P_2 are completely conflicting, $P_1(j) = 1,\ P_2(k) = 1$ for $j \neq$ k, then $cos(P_1, P_2) = 0$ and $\|P\|^2 = \frac{1}{4} + \frac{1}{4} + 0 = \frac{1}{2}$.	If π_1 and π_2 are two certain possibility distributions, then $\|\pi_1\| = 1$ and $\|\pi_2\| = 1$ and there are two cases of interest. 1) If π_1 and π_2 are completely compatible, $\pi_1(j) = \pi_2(j) = 1$ for some j. Here $cos(\pi_1, \pi_2) = 1$ and $\|\pi\|^2 = \frac{1}{4}(1) + \frac{1}{4}(1) + \frac{1}{2}(1) = 1$ 2) If π_1 and π_2 are completely conflicting, $\pi_1(j) = 1, \pi_2(k) = 1$ for $j \neq$ k, then $cos(\pi_1, \pi_2) = 0$ and $\|\pi\|^2 = \frac{1}{4} + \frac{1}{4} + 0 = \frac{1}{2}$.
If P_1 and P_2 are both maximally uncertain distributions, they have $P_1(j) = P_2(j) = \frac{1}{n}\ for\ all\ j$. Here $\|P_1\| = \|P_2\| = (1/n)^{1/2}$ and since we have shown in this case that $cos(P_1, P_2) = 1$ then $\|P\|^2 = \frac{1}{4}\frac{1}{n} + \frac{1}{4}\frac{1}{n} + \frac{1}{2}\left(\frac{1}{n}\right)^{1/2}\left(\frac{1}{n}\right)^{1/2} 1 = \frac{1}{n}$	If π_1 and π_2 are both maximally uncertain distributions, they have $\pi_1(j) = \pi_2(j) = 1$ for all j. Here $\|\pi_1\| = \|\pi_2\| = (n)^{1/2}$ and since we have shown in this case that $cos(\pi_1, \pi_2) = 1$ then $\|\pi\|^2 = \frac{1}{4}(n) + \frac{1}{4}(n) + \frac{1}{2}\left(n^{\frac{1}{2}}\right)\left(n^{\frac{1}{2}}\right) 1 = n$.

If P_1 is completely certain distribution, $\|P_1\| = 1$, and P_2 is completely uncertain distribution, $\|P_2\| = \left(\frac{1}{n}\right)^{1/2}$, then $P_1 P_2 = \frac{1}{n}$ and $\|P\|^2 = \frac{1}{4}(1) + \frac{1}{4}\frac{1}{n} + \frac{1}{2}\frac{1}{n} = \frac{3+n}{4n}$	If π_1 is completely certain distribution, $\|\pi_1\| = 1$, and π_2 is completely uncertain distribution, $\|\pi_2\| = (n)^{1/2}$, then $< \pi_1, \pi_2 > = 1$ and $\|\pi\|^2 = \frac{1}{4}(1) + \frac{1}{4}(n) + \frac{1}{2}(1) = \frac{1+n+2}{4} = \frac{3+n}{4}$.

Let us now consider the general case where we have *t* possibility distributions $\pi_1, \pi_2, \ldots, \pi_t$ where $\pi_i = [\pi_i(1), \pi_i(2), \ldots, \pi_i(n)\ ]$ with $\pi = \frac{1}{t}\sum_{i=1}^{t} \pi_i$ and where each component of π is being defined as: $\pi(j) = \frac{1}{t}\sum_{i=1}^{t} \pi_i(j)$.

Table 6. Comparative equations associated with the calculation of $\|\pi\|^2$ for *t* distributions

Probabilistic Framework	Possibilistic Framework
$\|P\|^2 = \sum_{j=1}^{n}\big(p(j)\big)^2 = \sum_{j=1}^{n}\left(\frac{1}{t^2}\right)\big(p_1(j) + p_2(j) + \ldots + p_i(j)\big)^2$	$\|\pi\|^2 = \sum_{j=1}^{n}\big(\pi(j)\big)^2 = \sum_{j=1}^{n}\left(\frac{1}{t^2}\right)\big(\pi_1(j) + \pi_2(j) + \ldots + \pi_i(j)\big)^2$
$\|P\|^2 = \sum_{i=1}^{n}\frac{1}{t^2}(\|P_i\|)^2 + \sum_{i=1}^{t}\sum_{\substack{k=1 \\ k\neq i}}^{t}\frac{1}{t^2} < P_i, P_k >$	$\|\pi\|^2 = \sum_{i=1}^{n}\frac{1}{t^2}(\|\pi_i\|)^2 + \sum_{i=1}^{t}\sum_{\substack{k=1 \\ k\neq i}}^{t}\frac{1}{t^2} < \pi_i, \pi_k >$
$\|P\|^2 = \sum_{i=1}^{n}\frac{1}{t^2}(\|P_i\|)^2 + \frac{2}{t^2}\sum_{i=1}^{t-1}\sum_{k=i+1}^{t} < P_i, P_k >$	$\|\pi\|^2 = \sum_{i=1}^{n}\frac{1}{t^2}(\|\pi_i\|)^2 + \frac{2}{t^2}\sum_{i=1}^{t-1}\sum_{k=i+1}^{t} < \pi_i, \pi_k >$

Let us take a look at the formula $\|\pi\|^2 = \sum_{i=1}^{n}\frac{1}{t^2}(\|\pi_i\|)^2 + \frac{2}{t^2}\sum_{i=1}^{t-1}\sum_{k=i+1}^{t} < \pi_i, \pi_k >$ and consider the situation where we have two categories of possibility distributions: *complete knowledge* and *total ignorance* distributions:

- One being a *certainty* distribution; only one of its components is one (*complete knowledge*).
- The other is a pure *uncertainty* distribution; here, all elements are 1 *(total ignorance*).

First, for π_i that has certainty, then $NegEnt = 1$. In contrast, for any π_i that is pure uncertainty, we have shown that $NegEnt = \frac{1}{n}$. Table 7 below gives results for the calculations of the dot product. Table 8 below presents the calculations of $NegEnt$ in a mixed case i.e. that t_1 of the distributions are pure certainty and t_2 are pure uncertainty.

Table 7. Comparative equations associated with the dot product under both frameworks

Probabilistic Framework	Possibilistic Framework
If both p_i and p_k represent **certainty** then if they agree on the certainty element $< p_i, p_k >= 1$ and if they disagree, then $< p_i, p_k > = 0$. If one of $< p_i, p_k >$ is pure uncertainty, for example p_i, then: - if the other distribution p_k is certainty, we have $< p_i, p_k >= \sum_{j=1}^{n} p_i(j)p_k(j) = 1\frac{1}{n} = \frac{1}{n}$. - if the other distribution is all pure uncertainty, we get $< p_i, p_k >= \sum_{j=1}^{n} p_i(j)p_k(j) = \frac{n}{n^2} = \frac{1}{n}$. - if one of $p_i$or p_k is pure uncertainty then $< p_i, p_k >= \frac{1}{n}$.	If both π_i and π_k are **certainty** then if they agree on the certainty element $< \pi_i, \pi_k > = 1$ and if they disagree $< \pi_i, \pi_k > = 0$. If one of $< \pi_i, \pi_k >$ is pure uncertainty, for example π_i, then: - if the other distribution π_k is certainty, we have $< \pi_i, \pi_k > = \sum_{j=1}^{n} \pi_i(j)\pi_k(j) = 1$. - if the other distribution is all pure uncertainty, we get $< \pi_i, \pi_k > = \sum_{j=1}^{n} \pi_i(j)\pi_k(j) = n$.. - if we have only one possibility distribution that is certain then $< \pi_i, \pi_k > = 1$ or - if all possibility distributions are certain and they agree on the certainty element then $< \pi_i, \pi_k > = 1$.

Table 8. Comparative equations associated with $\|\pi\|^2$ in mixed case under both frameworks

Probabilistic Framework	Possibilistic Framework
$\|P\|^2 = \frac{1}{t^2}\left[t_1 + \frac{1}{n}(t - t_1) + \frac{1}{n}((t - t_1)(t + t_1 - 1)) + \sum_{j=1}^{n} g_j(g_j - 1)\right]$ $\|P\|^2 = \frac{1}{t^2}\left[t_1 + \sum_{j=1}^{n} g_j(g_j - 1) + \frac{1}{n}((t - t_1)(t + t_1))\right]$ $\|P\|^2 = \frac{1}{t^2}\left[t_1 + \sum_{j=1}^{n} g_j(g_j - 1) + \frac{1}{n}(t^2 - t_1^2)\right]$ In the special case where all the pure certain distributions agree, $g_1 = t_1$ and all other $g_j = 0$, we get $$\|P\|^2 = \frac{1}{t^2}\left[t_1 + t_1(t_1 - 1) + \frac{1}{n}(t^2 - t_1^2)\right]$$ $$\|P\|^2 = \frac{1}{t^2}\left[t_1^2 + \frac{1}{n}(t^2 - t_1^2)\right]$$ $$NegEnt = \frac{1}{t^2}\left[t_1^2 + \frac{1}{n}(t^2 - t_1^2)\right]$$	$\|\pi\|^2 = \frac{1}{t^2}\left[t_1 + n(t - t_1) + n((t - t_1)(t + t_1 - 1)) + \sum_{j=1}^{n} g_j(g_j - 1)\right]$ $\|\pi\|^2 = \frac{1}{t^2}\left[t_1 + \sum_{j=1}^{n} g_j(g_j - 1) + n((t - t_1)(t + t_1))\right]$ $\|\pi\|^2 = \frac{1}{t^2}\left[t_1 + \sum_{j=1}^{n} g_j(g_j - 1) + n(t^2 - t_1^2)\right]$ In the special case where all the pure certain distributions agree, $g_1 = t_1$ and all other $g_j = 0$ we get $$\|\pi\|^2 = \frac{1}{t^2}[t_1 + t_1(t_1 - 1) + n(t^2 - t_1^2)]$$ $$\|\pi\|^2 = \frac{1}{t^2}[t_1^2 + n(t^2 - t_1^2)]$$ $$NegEnt = \frac{t^2}{[t_1^2 + n(t^2 - t_1^2)]}$$

6. Uniform fusion of possibility distribution

The fusion of multi-source probabilistic information has been detailed in Section 5 of Yager and Petry [1]. This section discusses the main differences between both approaches. Assume X is a variable that takes its value in the space $\Omega = \{x_1, x_2, \dots, x_t\}$. Let π_i be a possibility distribution on Ω representing the information provided by source i regarding the value of X so that, each $\pi_i = [\pi_i(1), \pi_i(2), \dots, \pi_i(n)]$ with t possibility distributions, π_i, for $i = 1\ to\ t$, then the basic uniform fusion of these distributions is a possibility distribution π as given in Table 9.

Table 9. Equations for the fusion of distributions for both frameworks

Probabilistic Framework	Possibilistic Framework
$P = \frac{1}{t}\sum_{i=1}^{t} P_i$ with $P(j) = \frac{1}{t}\sum_{i=1}^{t} P_i(j)$	$\pi = \frac{1}{t}\sum_{i=1}^{t} \pi_i$ with $\pi(j) = \frac{1}{t}\sum_{i=1}^{t} \pi_i(j)$
$\|P\|^2 = \frac{1}{t^2}\left[\sum_{i=1}^{t}\|P_i\|^2 + 2\sum_{i=1}^{t-1}\sum_{k=i+1}^{t} < P_i . P_k >\right]$	$\|\pi\|^2 = \frac{1}{t^2}\left[\sum_{i=1}^{t}\|\pi_i\|^2 + 2\sum_{i=1}^{t-1}\sum_{k=i+1}^{t} < \pi_i . \pi_k >\right]$
The total number of terms being combined: $t + \frac{2(t)(t-1)}{2} = t^2$ and $NegEnt = \|P\|^2 \in [0,1]$.	The total number of terms being combined: $t + \frac{2(t)(t-1)}{2} = t^2$ and $NegEnt \in \left[0, \frac{1}{n}\right]$
$AveConf(P_i, P_k) = \frac{2}{(t)(t-1)}\sum_{i=1}^{t-1}\sum_{k=i+1}^{t}\left(1 - \frac{P_i P_k}{\|P_i\|\|P_k\|}\right)$ $= 1 - \frac{2}{(t)(t-1)}\sum_{i=1}^{t-1}\sum_{k=i+1}^{t}\left(\frac{P_i P_k}{\|P_i\|\|P_k\|}\right)$	$AveConf(\pi_i, \pi_k) = \frac{2}{t(t-1)}\left[\sum_{i=1}^{t-1}\sum_{k=i+1}^{t}\left(1 - \frac{\pi_i \pi_k}{\|\pi_i\|\|\pi_k\|}\right)\right]$ $= 1 - \frac{2}{t(t-1)}\left[\sum_{i=1}^{t-1}\sum_{k=i+1}^{t}\left(\frac{\pi_i \pi_k}{\|\pi_i\|\|\pi_k\|}\right)\right]$

The objective of fusing t possibility distributions, $\pi_i, for\ i = 1\ to\ t$, provided by multiple sources, is to obtain a fused estimate that contains the largest amount of information about the value of X. That means to obtain a large value of $NegEnt$. As in the case of the probabilistic framework, pairing possibility distributions that are non-conflicting, i.e. a $\pi_i \pi_k$ large, tend to increase the information. Alternatively, those pairs with small compatibility, i.e. a $\pi_i \pi_k$ small, tend to increase $NegEnt$ as well. The explication is given in Yager and Petry [1] as that the former pairs affect the value $\left(1/t^2\right)$. We note that $t^2 = t + t(t-1)$ is the number of possibility distributions plus the number of pairs. Thus while a conflicting pair does not affect too much the sum $\sum_{i=1}^{t-1}\sum_{k=i+1}^{t}(\pi_i \pi_k)$, it can however augment the value of $NegEnt$ since it is counted in the term t^2.

Yager and Petry [1] point out that to only fuse the possibility distributions that have a high compatibility provokes a loss of credibility of the fusion. They suggest the use of a set measure that indicates the credibility of a fusion based on a simple weighted average of a subset of the distributions. This can be applied to possibility distributions as well.

The fusion process considers the possibility distribution subset B of the most credible distributions. Thus, the set measure $Cred: 2^Z \rightarrow [0,1]$ is defined such that for any subset B of Z, $Cred(B)$ indicates the credibility of a fusion based on only the distributions in B. The credibility of a possibility distribution subset requires the following natural properties:

1. $Cred(\emptyset) = 0$;
2. $Cred(Z) = 1$; and
3. $If\ A \subseteq B\ then\ Cred(A) \leq\ Cred(B)$.

The search and rescue example of Yager and Petry [1] that consists of three distributions, and which represent information on four spatial locations, is being reused here for the possibilistic case. The decision is about finding the location where to rescue? The sources of probabilistic information are, hypothetically, from UAVs, surveillance aircraft, and human sources. For the purpose of comparison let us state that the sources of possibilistic information on the same locations are from humans. Table 10 shows the comparison. As expected, it does not gives the same results since we use normalized distributions in the possibilistic framework. The choice of using a framework or another is based upon the nature and characteristics of the information sources. Consequently, the analysis presented in Section 5 of Yager & Petry on the use of $Conf$ and $NegEnt$ for the probabilistic framework also applies here to the possibilistic framework.

Table 10. Calculations of conflicts and information for the example given in Yager and Petry [1]

Probabilistic Framework	Possibilistic Framework
P_1: (.5, .2, .2, .1); P_2: (.4, .3, .2, .1); P_3: (.1, .2, .1, .6)	π_1: (.5, .2, .2, .1); Normalisation → π_1: (1, .2, .2, .1); π_2: (.4, .3, .2, .1); Normalisation → π_2: (1, .3, .2, .1); π_3: (.1, .2, .1, .6) Normalisation → π_3: (.1, .2, .1,1)
$\|P_1\| = \left(\sqrt{.34}\right) = .583;\ NegEnt = .34$ $\|P_2\| = \left(\sqrt{.3}\right) = .547;\ NegEnt = .3$ $\|P_3\| = \left(\sqrt{.42}\right) = .648;\ NegEnt = .42$ $P_1 \cdot P_2 = .31;\ P_1 \cdot P_3 = .17; P_2 \cdot P_3 = .18$	$\|\pi_1\| = \left(\sqrt{1.09}\right) = 1.0440;\ NegEnt$ =0.9174 $\|\pi_2\| = \left(\sqrt{1.14}\right) = 1.0677;\ NegEnt = 0.8772$ $\|\pi_3\| = \left(\sqrt{1.06}\right) = 1.0296;\ NegEnt = 0.9434$ $\pi_1 \cdot \pi_2 = 1.11;\ \pi_1 \cdot \pi_3 = .26; \pi_2 \cdot \pi_3 = .28$
$\text{Conf}(P_1, P_2) = 1 - \frac{.31}{.583 * .547} = 1 - \frac{.31}{.318}$ $= 1 - .975 = .025$ $\text{Conf}(P_1, P_3) = 1 - \frac{.17}{.583 * .648} = 1 - \frac{.17}{.378}$ $= 1 - .450 = .550$	$\text{Conf}(\pi_1, \pi_2) = 1 - \frac{1.11}{1.0440 * 1.0677}$ $= 1 - \frac{1.11}{1.11478} = 1 - 0.9958$ $= 0.0042$

$\text{Conf}(P_2, P_3) = 1 - \frac{.18}{.547 * .648} = 1 - \frac{.18}{.354}$ $= 1 - .508 = .452$	$\text{Conf}(\pi_1, \pi_3) = 1 - \frac{.26}{1.0440 * 1.0296}$ $= 1 - \frac{.26}{1.0749} = 1 - 0.2419$ $= 0.7581$ $\text{Conf}(\pi_2, \pi_3) = 1 - \frac{.28}{1.0296 * 1.0296}$ $= 1 - \frac{.28}{1.0993} = 1 - 0.2547$ $= 0.7453$
$\|R(P_1, P_2)\|^2 = \frac{.316}{.32} = .988$ $\|R(P_1, P_3)\|^2 = \frac{.275}{.38} = .724$ $\|R(P_2, P_3)\|^2 = \frac{.27}{.36} = .75$ $R(P_1, P_2, P_3) = (.333, .233, .166, .266);$ $\|R(P_1, P_2, P_3)\|^2 = .264$	$\pi(\pi_1, \pi_2) = (1; .25; .2; .1)$ $NegEnt(\pi(\pi_1, \pi_2)) = .8989$ $\pi(\pi_1, \pi_3) = (1; .2; .15; 1)$ $NegEnt(\pi(\pi_1, \pi_3)) = .4848$ $\pi(\pi_2, \pi_3) = (1; .25; .15; 1)$ $NegEnt(\pi(\pi_2, \pi_3)) = .4796$ $\pi(\pi_1, \pi_2, \pi_3) = (1; .233; .166; .4)$ $NegEnt(\pi(\pi_1, \pi_2, \pi_3)) = .8050$

7. On weighted average fusion

The previous section shows how to use the distributions and their information content as an initial step for choices of possibility distributions. In this section, we discuss how to use the credibility, $Cred$, along with $NegEnt$ for the final fusion taking into account the credibility of fused values using different subsets of $Z = \{\pi_1, \pi_2, \ldots, \pi_t\}$. Given a subset B of possibility distributions from Z we can calculate the associated fused value, π_B. In particular if $|B|$ is the number of distributions in B then $\pi_B = \frac{1}{|B|}\sum_{\pi_i \in B} \pi_i$, and $\|\pi_B\|^2 \in [0,1]$ is given by:

$$\|\pi_B\|^2 = \frac{1}{|B|^2}\left[\sum_{\pi_i \in B} \|\pi_i\|^2 + 2\sum_{\substack{i=1 \\ \pi_i \in B}}^{t-1} \sum_{\substack{k=i+1 \\ \pi_k \in B}}^{t} \pi_i \pi_k\right] \tag{12}$$

The measure, $Cred$, provides the credibility associated with π_B and its range is given by $Cred(B) \in [0,1]$. The decision problem is now to find a fused value that has both high values for $Cred(B)$ and for $NegEnt(\pi_B)$. To select which subset to use, authors in [1] use the dominance concept related to the credibility and quantity of information. It is expressed as the predicate $Dom(B_i, B_j)$:

$$Dom(B_i, B_j) = \left[\left(Cred(B_i) \geq (Cred(B_j)\right) \wedge (NegEnt \geq NegEnt) \wedge Cond(>, <)\right] \tag{13}$$

Where $Cond(>)$ is true only if at least one of the "≥"is ">. If $Dom(B_i, B_j)$ is true then B_i dominates B_j and so B_j can be removed from consideration. Based on dominance rules, we have a collection of non-dominated

fusions where each fusion is determined from a subset of space Z denoted as $B_1, \ldots, B_r$. For each subset B_j, we can calculate its associated fusion π_{B_j}, also its $NegEnt(B_j)$ value, its $NegEnt(\pi_{B_j})$ and its credibility, $Cred(B_j)$. The information about $NegEnt(B)$ and credibility values are being used to select among these possible fusions, the π_{B_j}. In the following, we consider how to choose the final possibility distribution based both on the information quantity and on credibility using the example of the previous section.

The collection of relevant possibility distributions is $Z = \{\pi_1, \pi_2, \pi_3\}$. So there are seven subsets of Z to consider: $B_1 = \{\pi_1\}$, $B_2 = \{\pi_2\}$, $B_3 = \{\pi_3\}$, $B_4 = \{\pi_1, \pi_2\}$, $B_5 = \{\pi_1, \pi_3\}$, $B_6 = \{\pi_2, \pi_3\}$, $B_7 = \{\pi_1, \pi_2, \pi_3\}$. In Table 11, we compare credibility functions proposed in [1] for the probabilistic case with the credibility function obtained for the possibilistic approach.

Table 11. Credibility functions used for both frameworks

<table>
<tr><th>Probabilistic Framework</th><th>Possibilistic Framework</th></tr>
<tr><td><u>First credibility function:</u> at least two distributions included in a subset is used.
$C1(B_1) = C1(B_2) = C1(B_3) = 0$
$C1(B_4) = C1(B_5) = C1(B_6) = C1(B_7) = 1$</td><td>- Based on correlation matrix $corr(\pi_1, \pi_2)$.
$Cred_i = \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{n} corr(\pi_i, \pi_j)$ with
$corr(\pi_1, \pi_2) = \frac{<\pi_1, \pi_2>}{\|\pi_1\|.\|\pi_2\|}$</td></tr>
<tr><td><u>Second credibility function:</u> “most” as the fuzzy value M
$M = \begin{cases} 0 & |B| = 1 \\ .7 & |B| = 2 \\ 1.0 & |B| = 3 \end{cases}$
$C2(B_1) = C2(B_2) = C2(B_3) = 0$;</td><td>N/A</td></tr>
<tr><td><u>Dom -assessment</u>
<table>
<tr><th>$Subset$</th><th>$NegEnt$</th><th>$C1$</th><th>$C2$</th></tr>
<tr><td>B_1</td><td>.34</td><td>0</td><td>.0</td></tr>
<tr><td>B_2</td><td>.3</td><td>0</td><td>.0</td></tr>
<tr><td>B_3</td><td>.42</td><td>0</td><td>.0</td></tr>
<tr><td>B_4</td><td>.315</td><td>1</td><td>.7</td></tr>
<tr><td>B_5</td><td>.275</td><td>1</td><td>.7</td></tr>
<tr><td>B_6</td><td>.27</td><td>1</td><td>.7</td></tr>
<tr><td>B_7</td><td>.264</td><td>1</td><td>1</td></tr>
</table></td><td><u>Dom -assessment</u>
<table>
<tr><th>$Subset$</th><th>$NegEnt$</th><th>$Cred$</th></tr>
<tr><td>B_1</td><td>.9174</td><td>.7937</td></tr>
<tr><td>B_2</td><td>.8772</td><td>.7946</td></tr>
<tr><td>B_3</td><td>.9434</td><td>.4668</td></tr>
<tr><td>B_4</td><td>.8989</td><td>.7952</td></tr>
<tr><td>B_5</td><td>.4848</td><td>.8392</td></tr>
<tr><td>B_6</td><td>.4796</td><td>.8420</td></tr>
<tr><td>B_7</td><td>.8050</td><td>.8697</td></tr>
</table></td></tr>
<tr><td><u>Results :</u>
$C1$: non-dominated subsets: $ND = \{B_3, B_4\}$
$C2$: non-dominated subsets: $ND = \{B_1, B_4, B_7\}$</td><td><u>Results:</u>
$ND = \{B_1, B_3, B_4, B_7\}$.
$\{B_2, B_5, B_6\} \notin ND$.</td></tr>
</table>

The dominance principle (Eq.13) is applied to get the results of Table 11. The results obtained with the possibilistic approach use the follow statements:

- $NegEnt(B_4) > NegEnt(B_2),\ NegEnt(B_7) > NegEnt(B_5), NegEnt(B_7) > NegEnt(B_6)$
- $\big(Cred(B_4) > Cred(B_2), Cred(B_7) > Cred(B_5), Cred(B_7) > Cred(B_6)\big)$
- B_1, B_3, B_4, B_7 dominate all other subsets B_2, B_5, B_6 ;

A quick look at the results obtained using Yager & Petry [1] credibility functions as compared to the possibilistic approach shows that the most best subsets are selected using the possibilistic framework. Yager and Petry [1] propose another approach for obtaining the quality of a fusion by the use of prioritized aggregation introduced by Yager [23] where a score associated with a fusion is a weighted sum of its credibility and $NegEnt$. In our case, we start with the set $Z = \{\pi_1, \pi_2, \ldots, \pi_t\}$ and choose a subset B_i of Z. The fusion is performed by taking an equally weighted aggregation of the possibility distributions in B_i.
In [1], the authors suggest a more general approach based on a non-uniform weighted average of the elements in Z. Here, we propose a new weighting approach. The weighted factor of a possibility distribution is denoted by w, which is determined by both the credibility $Cred$ and the separability *SE*. The weighting factor, *w*, is defined as:

$$w = \frac{1}{2}\big(Cred + Cred \cdot SE^{-Cred}\big) \tag{14}$$

The factor $1/2$ is used to normalized w and to guarantee that $0 \leq w \leq 1$. The separability degree of two possibility distributions is obtained based on Manhattan distance as:

$$SE = 1 - \left(\frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \big(\pi_i - \pi_j\big) \right) \tag{15}$$

Finally, the functional processes for an intelligent quality-based approach to fuse multi-source possibilistic information can be summarized as shown in Figure 5 as an ‘8-step’ methodology. This methodology is being used, in the next section, to conduct an empirical comparison between both frameworks.

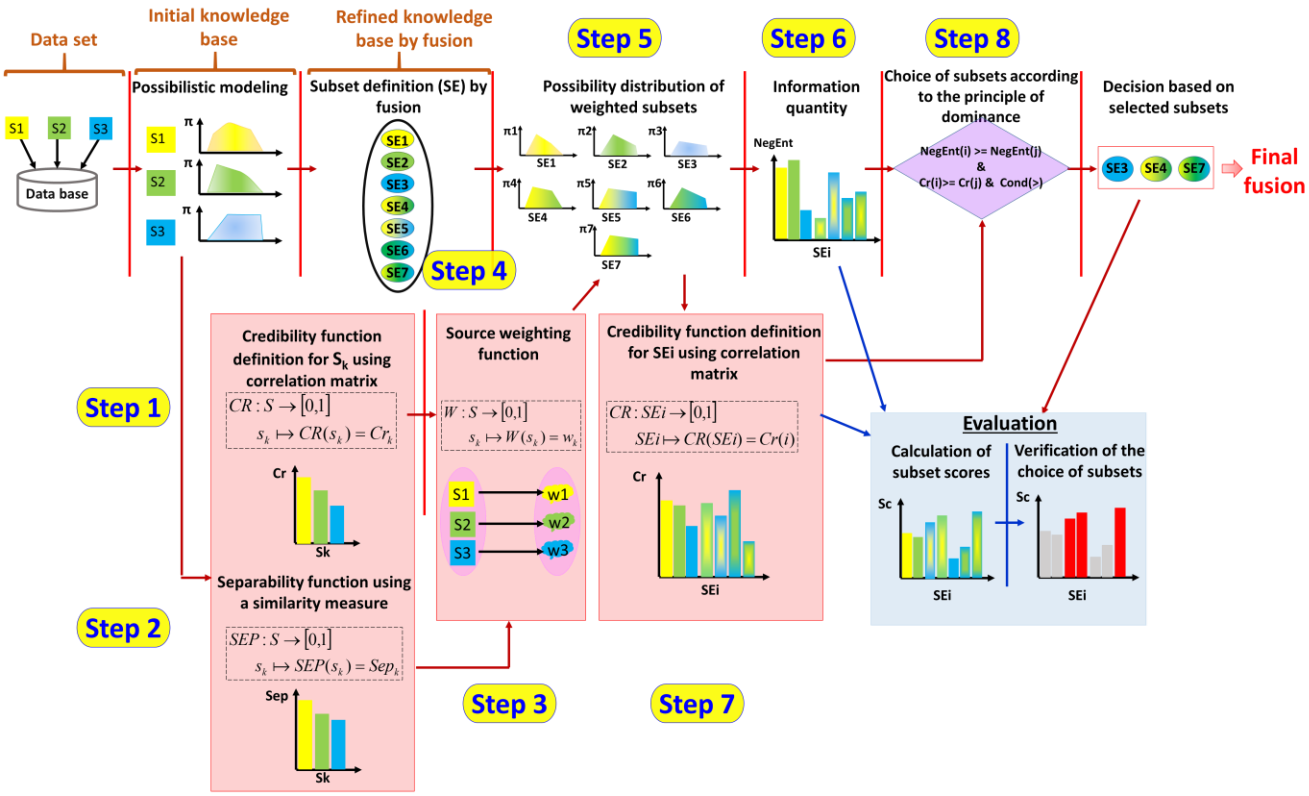


Figure 5. The functional processing (8-step methodology) of the possibilistic approach

8. Empirical results

This section details the calculations of the possibilistic approach to get the results presented in the right column of Table 11 as well as to illustrate the functional processing of Figure 5. A numerical example is shown first. The numerical example is the same than the one presented in the previous section. The second part shows an experimentation with four benchmark experimental datasets: IRIS Fisher data set [24], Pima Indians Diabetes [25, 26], Glass data set [27] and Liver-disorders [28] that have different degrees of information quality namely about information incompleteness.

8.1 Numerical example

The set Z, of possibility distributions are considered as the sources of information, $Z = \{\pi_1, \pi_2, \pi_3\}$, where π_1: (.5,.2,.2,.1); π_2: (.4,.3,.2,.1); π_3: (.1,.2,.1,.6). In general, for the selection of distributions, if a possibility distribution has a small conflict with respect to other distributions, then, that distribution should contribute more on the final fusion results than the other distributions that show high conflicts. A weighting function must then be defined to reflect that. To determine the value of each weight, a correlation coefficient between two possibility distributions is used that leads to the definition of a credibility measure for each distribution.

Using the same numerical example of the previous section, the 8-step methodology, illustrated in Figure 5, is unfolded the following way:

<u>Step 1:</u> Computation of the credibility measure for each possibility distribution based on one single source of knowledge

The credibility ($Cred$) of possibility distribution represents the similarity among distributions. The credibility values for the numerical example are computed accordingly to Table 11 and are respectively:

$$Cred(\pi_1) = 0.7103; Cred(\pi_2) = 0.7389; Cred(\pi_3) = 0.4785$$

The credibility measures show that the distributions, π_1 and π_2, are the most credible.

<u>Step 2:</u> Computation of the separability measure for each possibility distribution

This step is to define weights for each possibility distribution based on the distribution credibility and its discriminating power that is represented by the separability measure.

For the search and rescue example, the values of each degree of separability are:

$$SE(\pi_1) = 0.8333; SE(\pi_2) = 0.8667; SE(\pi_3) = 0.7667$$

<u>Step 3:</u> Computation of the weights for possibility distributions

The determination of weights for the distributions is related to the credibility $Cred$ and the separability measure *SE* which are two information quality-based factors. If both factors can be taken into consideration, then a better selection of distribution subsets will result. The credibility and separability weights can be derived based on possibility distributions alone, requiring no extra a priori knowledge. The weight factor of a possibility distribution is denoted by w, which is determined by both credibility $Cred$ and separability SE. It is defined by: $w = \frac{1}{2}\left(Cred + Cred \cdot SE^{-Cred}\right)$. The factor $1/2$ is used to normalize w and to guarantee that $0 \leq w \leq 1$. In our example, the weights are as follow:

$$w_1(\pi_1) = 0.7594; w_2(\pi_2) = 0.7801; w_3(\pi_3) = 0.5109$$

<u>Step 4:</u> Definition of all subsets of the distribution set Z

The example consists of 3 information sources, consequently 7 subsets are being constructed:

$$\begin{array}{c} Z \\ \hline B_1 = \{f_1\} \\ B_2 = \{f_2\} \\ B_3 = \{f_3\} \end{array} \left| \begin{array}{l} B_4 = \{f_1, f_2\} \\ B_5 = \{f_1, f_3\} \\ B_6 = \{f_2, f_3\} \end{array} \right| B_7 = \{f_1, f_2, f_3\} \quad (16)$$

<u>Step 5:</u> Fusion of the weighted possibility distributions of each subset

The corresponding distributions of the different subsets ($B_{k=1\ to\ 7}$) are given as follow.

$$\pi_{B_k}(\pi_i, \pi_j) = (w_i.\pi_i + w_j.\pi_j)/2 \quad (17)$$

$$B_1 = w_1.\pi_1 = \{0.3797, 0.1519, 0.1519, 0.0759\}$$
$$B_2 = w_2.\pi_2 = \{0.3120, 0.2340, 0.1560, 0.0780\}$$
$$B_3 = w_3.\pi_3 = \{0.0511, 0.1022, 0.0511, 0.3065\}$$
$$B_4 = (w_1.\pi_1 + w_2.\pi_2)/2 = \{0.3459, 0.1929, 0.1539, 0.0770\}$$
$$B_5 = (w_1.\pi_1 + w_3.\pi_3)/2 = \{0.2154, 0.1270, 0.1015, 0.1912\}$$
$$B_6 = (w_2.\pi_2 + w_3.\pi_3)/2 = \{0.1816, 0.1681, 0.1036, 0.1923\}$$
$$B_7 = (w_1.\pi_1 + w_2.\pi_2 + w_3.\pi_3)/3 = \{0.2476, 0.1627, 0.1197, 0.1535\}$$

Step 6: Computation of the information quantity ($NegEnt$) in each subset ($B_{k=1\ to\ 7}$)

This step performs the evaluation of the quality of each subset and selects only subsets that have a large degree of $NegEnt$. The information quantity ($NegEnt$) of each weighted subset is presented in Figure 6.

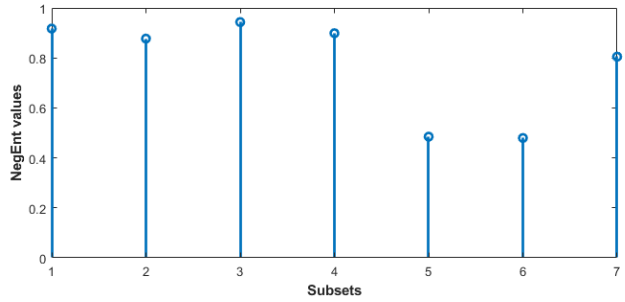


Figure 6. Information quantity ($NegEnt$) in each weighted subset

We remark the following:

1. The subset B_i with the largest $NegEnt$ provides the most information (in Figure 6, B_1 and B_3). For example, the subset B_1 gives a large $NegEnt$ value since B_4 fuses π_1 and π_2 that the first distribution (π_1) and the second distribution (π_2) have a small value of conflict.
2. Fusing a subset B_i and B_j with a large value of conflict provides a subset with a small $NegEnt$. For example, the fusion of π_1 (represented by B_1) or π_2 (represented by B_2) with π_3 (represented by B_3) gives a small value of $NegEnt$ as shown by B_5 , B_6.

Step 7: Computation of the credibility measure for each subset ($B_{k=1\ to\ 7}$)

Figure 7 provides the information about the credibility ($Cred$) of fused subsets, ($B_{k=1\ to\ 7}$). We can conclude based on this figure that the subset as B_5, B_6 and B_7 containing at least one credible distribution may have a high credibility value.

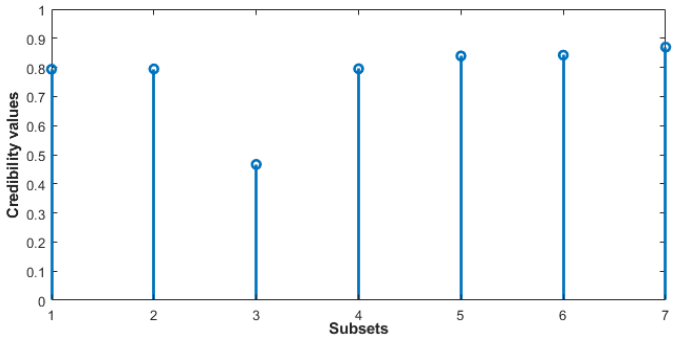


Figure 7. Credibility values of each subset

<u>Step 8:</u> Selection of the best subsets for the final fusion

Figure 8 presents the selected subsets to be considered in the final fusion. The selection process is based on the dominance concept that uses credibility, (*Cred*), and negative Entropy ($NegEnt$) values. The obtained results confirm that all subsets have a good information quantity except the subsets B_2, B_5, B_6.

$$\begin{array}{l|l|l|l|l|l|l|l} Subset & B_1 & B_2 & B_3 & B_4 & B_5 & B_6 & B_7 \\ Cred & 0.9174 & 0.8772 & 0.9434 & 0.8989 & 0.4848 & 0.4796 & 0.8050 \\ NegEnt & 0.7937 & 0.7946 & 0.4668 & 0.7952 & 0.8392 & 0.8420 & 0.8697 \end{array}$$

$\Downarrow$

$non - dominated\ subsets: ND = \{B_1, B_3, B_4, B_7\}$

Figure 8. Subsets selection

To verify the quality of a subset, it is possible to calculate a score associated with a fusion that is a weighted sum of its credibility and its information quantity ($NegEnt$). More formally the score of the fusion based on the subset B is defined as:

$$Score(B_k) = w_1 Cred(B_k) + w_2 NegEnt \tag{18}$$

$$\frac{w_1}{w_2} = \frac{Cred(B_k)}{1} \tag{19}$$

Normalizing the weights so that $w_1 + w_2 = 1$, we get

$$Score(B_k) = \frac{Cred(B_k)}{1 + Cred(B_k)}(1 + NegEnt) \tag{20}$$

The score, associated with each subset, is illustrated in Figure 9. Subsets selected according to the dominance rule are represented by the blue bars. Subsets with the highest scores are being selected according to the dominance principle.

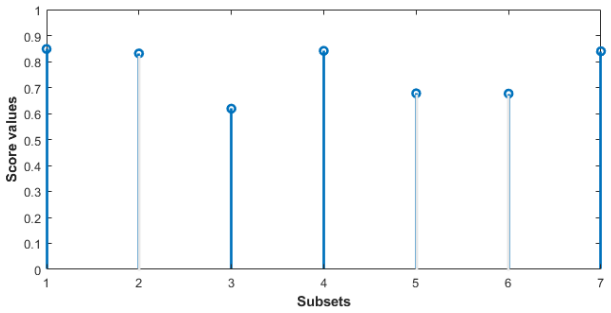


Figure 9. Score values of all subsets

The complexity of the proposed method depends on the complexity of the credibility measure, the complexity of the separability and the complexity of the information quantity measure. The complexity of these concepts is of :

- $O(log_2(t))$, while processed for each source of the *t* sources of knowledge.
- $O(2^t)$, while processed for all source of knowledge subsets.

The resulting complexity is equal to the maximum of [$O(log_2(t))$, $O(2^t)$] that corresponds to $O(2^t)$.

8.2 Experimental datasets as benchmark

This section presents a comparison between both approaches (probabilistic, possibilistic) following the 8-steps methodology of Figure 5 but this time on real dataset benchmarks. Four dataset benchmarks are being used: IRIS Fisher data set [24], Pima Indians Diabetes [25, 26], Glass data set [27] and Liver-disorders [28].

The Iris dataset consists of 3 classes, 4 attributes and 150 samples, with 50 samples for each class. In this paper, the first 35 samples of each class have been used for training. Each class C_i is being represented by the set $Z_i = \{\pi_{1,i}, \pi_{2,i}, \pi_{3,i}, \pi_{4,i}\}$ where $\pi_{j,i}$ is the distribution corresponding to a feature f_j.

The Pima Indians Diabetes is a dataset consisting of 2 classes, 8 attributes and 768 samples, with 500 samples for the first class and 268 samples for the second one. In this paper, the last 100 samples of the first class and the last 50 samples of the second class have been used for training. For the Pima Indians Diabetes database, each class C_i is represented by the set $Z_i = \{\pi_{1,i}, \pi_{2,i}, \pi_{3,i}, \pi_{4,i}, \pi_{5,i}, \pi_{6,i}, \pi_{7,i}, \pi_{8,i}\}$ where $\pi_{j,i}$ is the distribution corresponding to the feature f_j.

The Glass dataset consists of 2 classes, 9 attributes and 214 samples, where 163 samples are for the first class and 51 samples for the second one. In this paper, 58 samples of the first class and 15 samples of the second class are used for training. Using the Glass database, each class C_i is represented by the set $Z_i = \{\pi_{1,i}, \pi_{2,i}, \pi_{3,i}, \pi_{4,i}, \pi_{5,i}, \pi_{6,i}, \pi_{7,i}, \pi_{8,i}, \pi_{9,i}\}$ where $\pi_{j,i}$ is the distribution corresponding to feature f_j.

The Liver-disorders is a dataset consisting of 2 classes, 6 attributes and 326 samples, whose 138 samples for the first class and 188 samples for the second one. Using the Liver-disorders database, each class C_i is represented by the set $Z_i = \{\pi_{1,i}, \pi_{2,i}, \pi_{3,i}, \pi_{4,i}, \pi_{5,i}, \pi_{6,i}, \pi_{7,i}, \pi_{8,i}\}$ where $\pi_{j,i}$ is the distribution corresponding to feature f_j.

8.2.1 Probabilistic and possibilistic modeling of the information sources

Prior to apply the 8-step methodology of Figure 5, the information sources have to be modelled either by possibility or probability distributions. The construction of distributions is illustrated in Figure 10 for the IRIS data set for both frameworks: four features and three classes (color curves). The estimation process of probability distribution $p_{j,i}$ from feature f_j uses a histogram method. Then, the possibility distribution $\pi_{j,i}$ is inferred from the probability distribution based on the approach of Dubois and Prade [30].

For both probability and possibility modeling, the resolution of the histogram data that corresponds to the number of points n (bins) selected to represent the distributions is crucial since it directly impacts on the distribution's shape. For example, in Figure 10 the number of bins used to represent the distributions is 10. The total number of available samples is determinant in the choice of the number of bins in the histogram to represent classes according to features. Indeed, with a small amount of data, a large number of bins is not possible since the modeling of information associated with the feature becomes incorrect. The question now is what would be the adequate number of points to represent a feature?

To answer this question, a method based on the computation of the Shapley Index [31] is proposed in this study. The Shapley index is generally used to assess the quality of the feature discrimination for the resulting models. The information associated to a class according to feature values is modeled according to five different numbers of points n selected to quantify feature values, namely $n = 5$, 8, 10, 12 and 15 points. The Shapley Index is then calculated corresponding to each number of n. Figure 11 shows the results of the Shapley class discrimination values on the Iris dataset for the possibility modeling. In [32], the authors have demonstrated that the feature with the highest Shapley value is the one that has the largest discriminating factor between classes. Figure 11 shows that the highest Shapley value is obtained for feature 4 as $n = 10$. So, the number of points chosen to represent the probability and the possibility distributions has been fixed to $n = 10$.

With this value of $n = 10$, a comparison of a possibility class and a probability class is illustrated in Figure 10. It shows that the possibilistic framework gives distributions that have a greater potential to discriminate between classes (e.g. feature 2 and feature 4) than for a probabilistic framework.

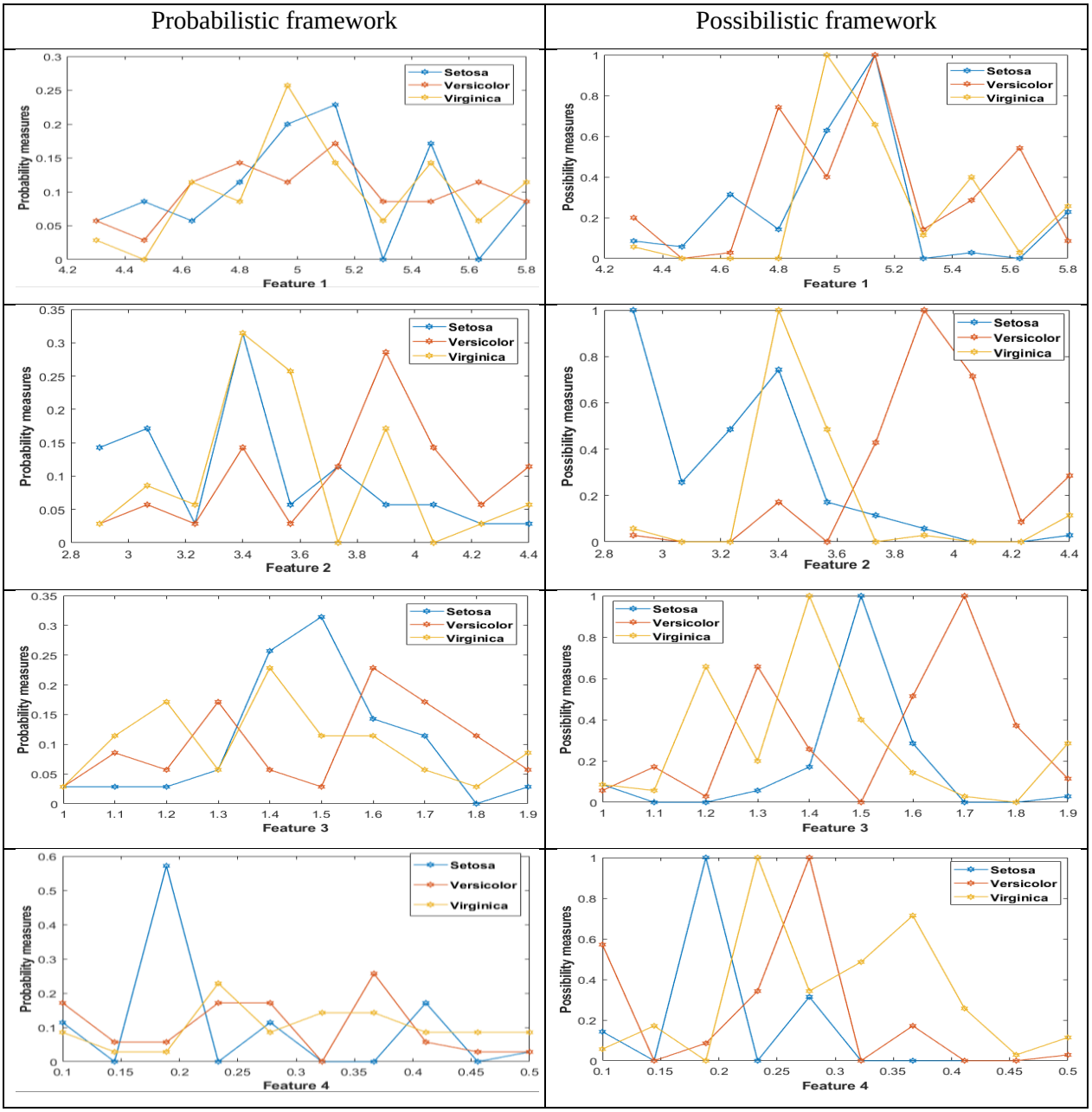


Figure 10. Class representations in probabilistic and possibilistic frameworks using the Iris dataset

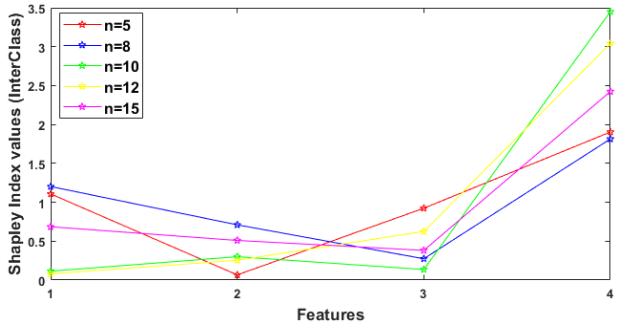


Figure 11. Shapley index values for class discrimination for the Iris dataset for $\boldsymbol{n = 5, 8, 10, 12}$ and $\boldsymbol{15}$

8.2.2 Comparison of the 8-step Probabilistic-Possibilistic methodology on real cases

The 8-step methodology of Figure 5 is now being compared but on real cases for both: the possibilistic and the probabilistic frameworks. The 8-steps methodology is listed as:

Step 1: Computation of the credibility measure for each distribution

Step 2: Computation of the separability measure for each distribution

Step 3: Computation of the weights for distributions

Step 4: Definition of all subsets of each set Z_i, i=1 to number of classes

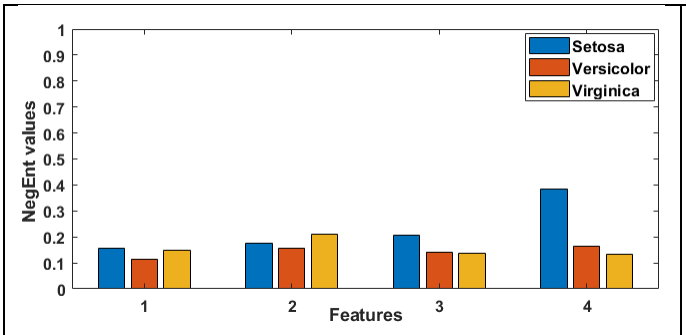
Step 5: Initial fusion of the weighted distributions of each subset

Step 6: Computation of the information quantity ($NegEnt$) in each subset $(B_{i,k},\ k=1\ to\ 2^m,\ m=number\ of\ features)$

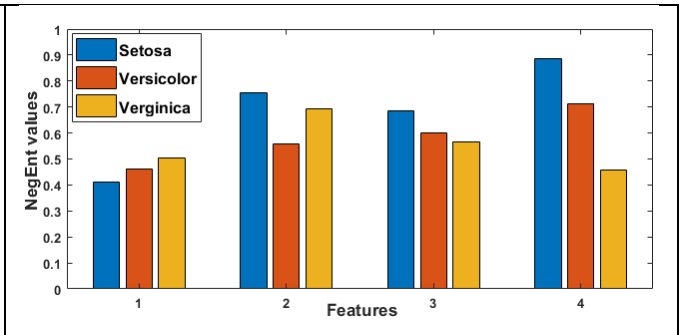
Step 7: Computation of a credibility measure for each subset $(B_{i,k})$

Step 8: Selection of the best subsets for the final fusion for each set Z_i

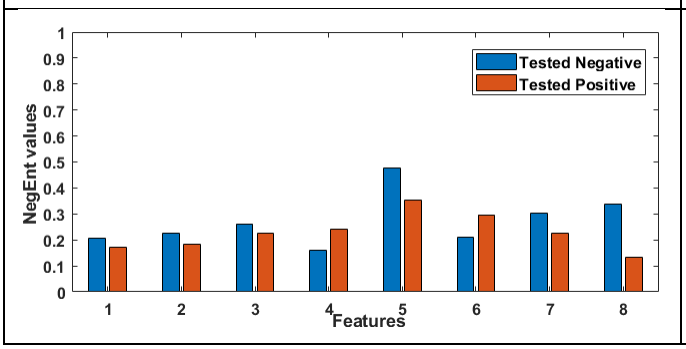
Before applying the methodology, let us give a primary idea of the dataset information quality. The information quantity ($NegEnt$) has been calculated for all distributions. Results are tabulated in Figure 12. The information quantity ($NegEnt$) obtained with the possibilistic framework gives larger values than with the probabilistic framework. For the four benchmarks, and based only on $NegEnt$, the possibilistic framework seems to offer a better modelling of the information sources and its imperfections than the probabilistic one.

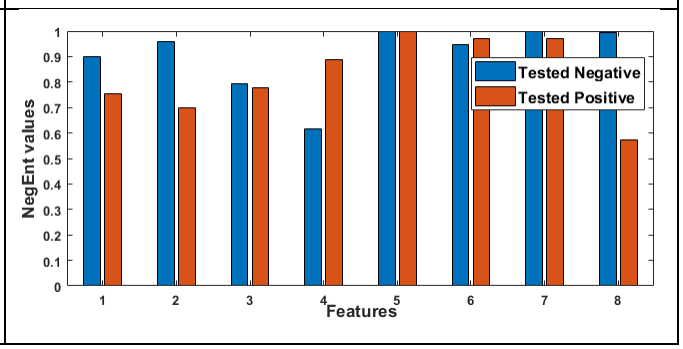


(a, Iris dataset)



(b, Iris dataset)





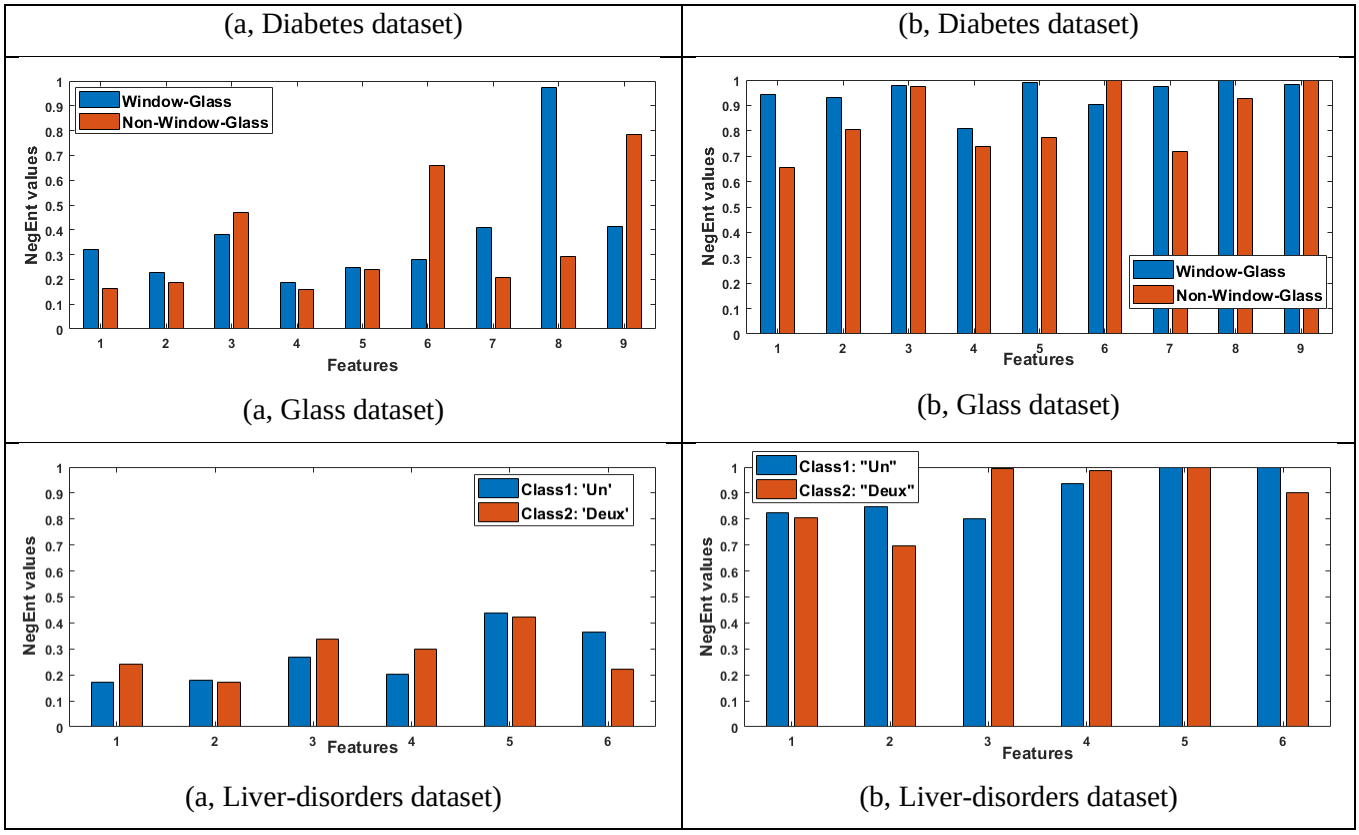
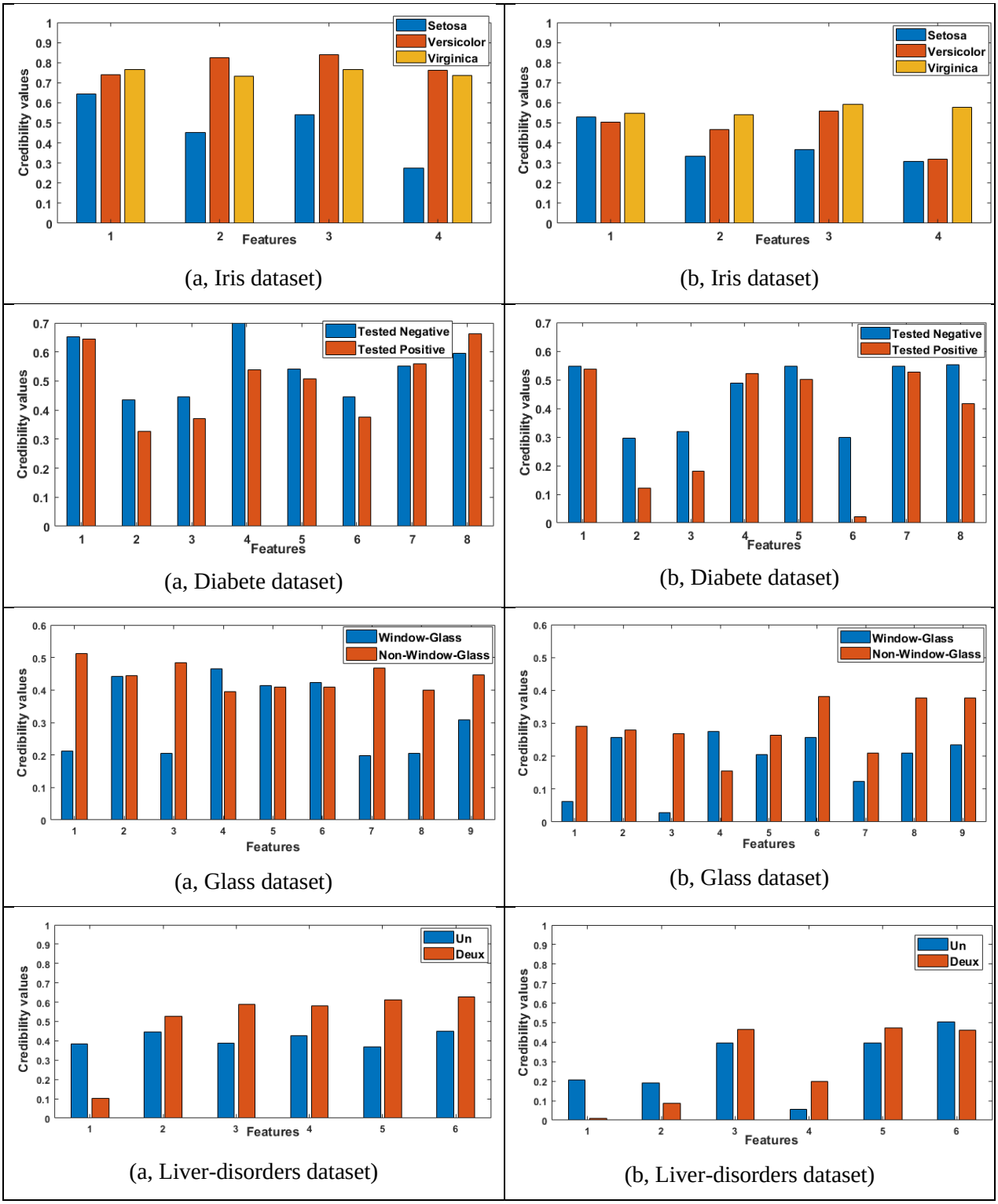


Figure 12. Information quantities (***NegEnt***) in (a) probabilistic framework and (b) possibilistic framework

Let us now apply the methodology according to both frameworks.

<u>Step 1:</u> Computation of the credibility measure for each distribution

The credibility degrees obtained with the possibilistic and the probabilistic frameworks are presented in Figure 13. We note different behaviors for credibility values in probabilistic and possibilistic frameworks. The credibility is defined as an average correlation value. The mean value is a linear operation. However, the correlation is calculated based on the cosine of the angle between the distributions. It is known that the cosine is a non-linear operation. So that the correlation is nonlinear. That nonlinearity is being observed, particularly for feature 4 of Iris dataset, for the variation of the credibility measures between the possibilistic and the probabilistic framework. For instance, in feature 4 of Iris dataset, the difference of credibility degrees between the first and the second class is very important for the probabilistic framework while we cannot see the corresponding behavior with the possibilistic framework. Other features like feature 6 of the Diabetes dataset and feature 6 of the Glass dataset present the same behavior.



(a, Iris dataset)

(b, Iris dataset)

(a, Diabete dataset)

(b, Diabete dataset)

(a, Glass dataset)

(b, Glass dataset)

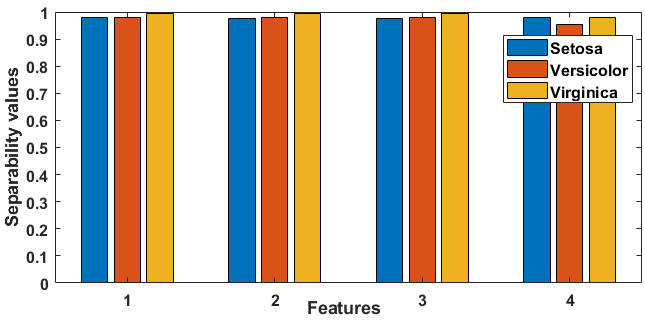
(a, Liver-disorders dataset)

(b, Liver-disorders dataset)

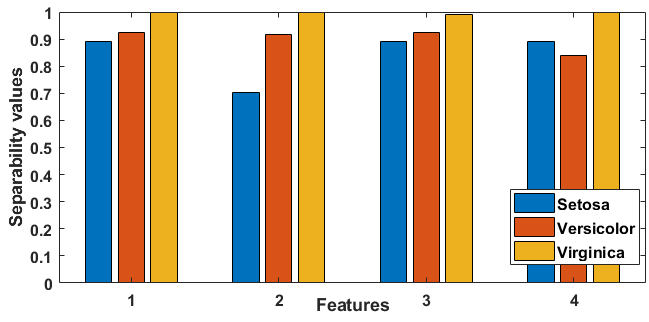
Figure 13. Credibility degrees in (a) probabilistic framework and (b) possibilistic framework

<u>Step 2:</u> Computation of the separability measure for each distribution

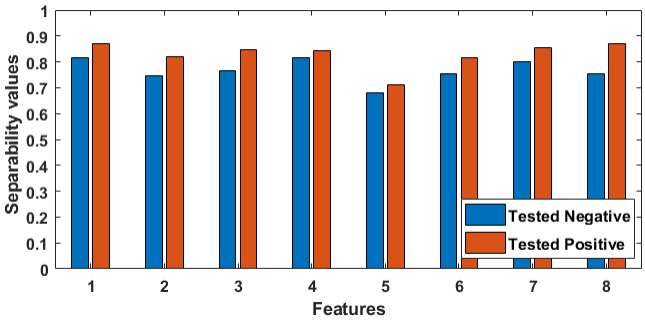
The separability degrees obtained in the possibilistic framework and the probabilistic framework are presented in Figure 14. This operation is linear, so it presents similar behavior for both frameworks. Figure 14 shows that the separability degree between classes for all features is very close to each other in probability framework. However, we observe a much better degree of separability with the possibilistic framework, which allows us to better distinguish between classes.



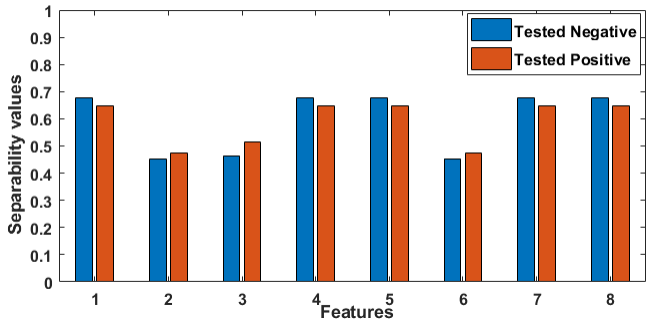
(a, Iris dataset)



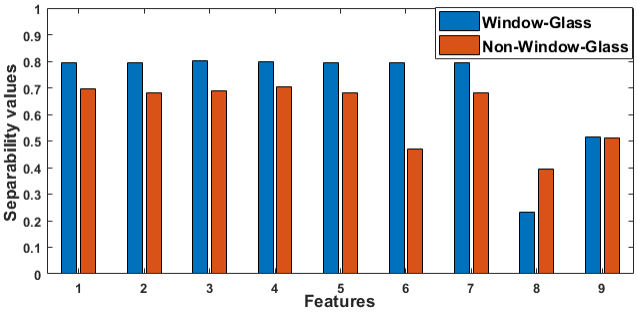
(b, Iris dataset)



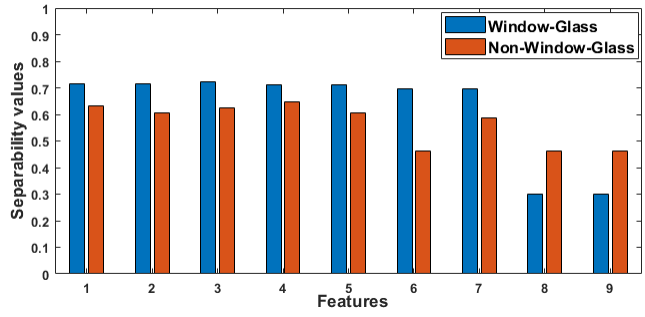
(a, Diabete dataset)



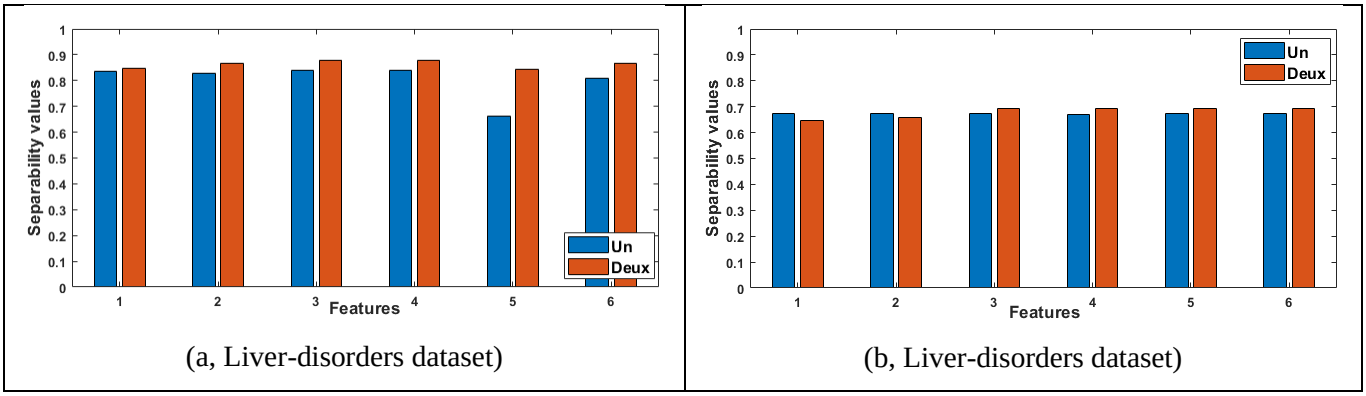
(b, Diabete dataset)



(a, Glass dataset)



(b, Glass dataset)



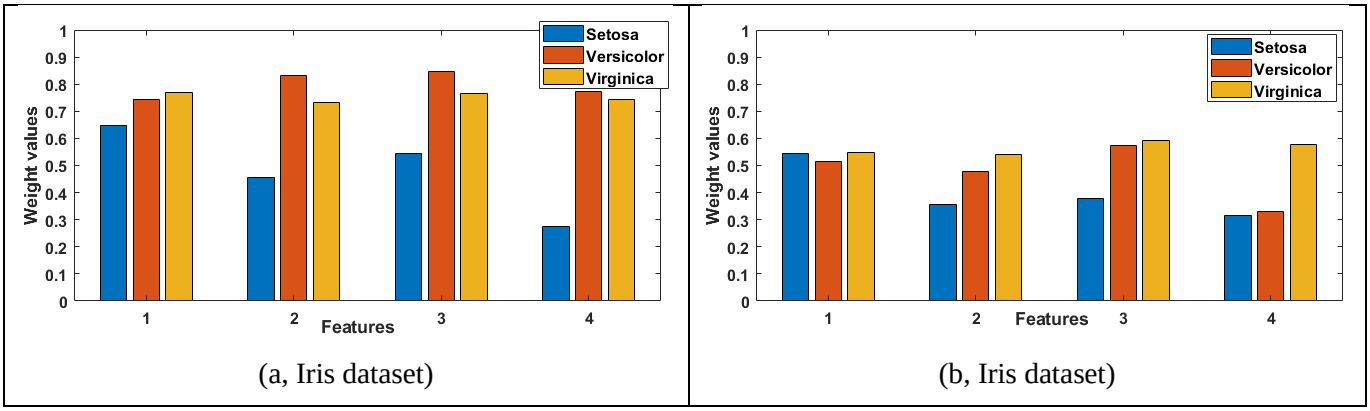
(a, Liver-disorders dataset) (b, Liver-disorders dataset)

Figure 14. Separability degrees in (a) probabilistic framework and (b) possibilistic framework

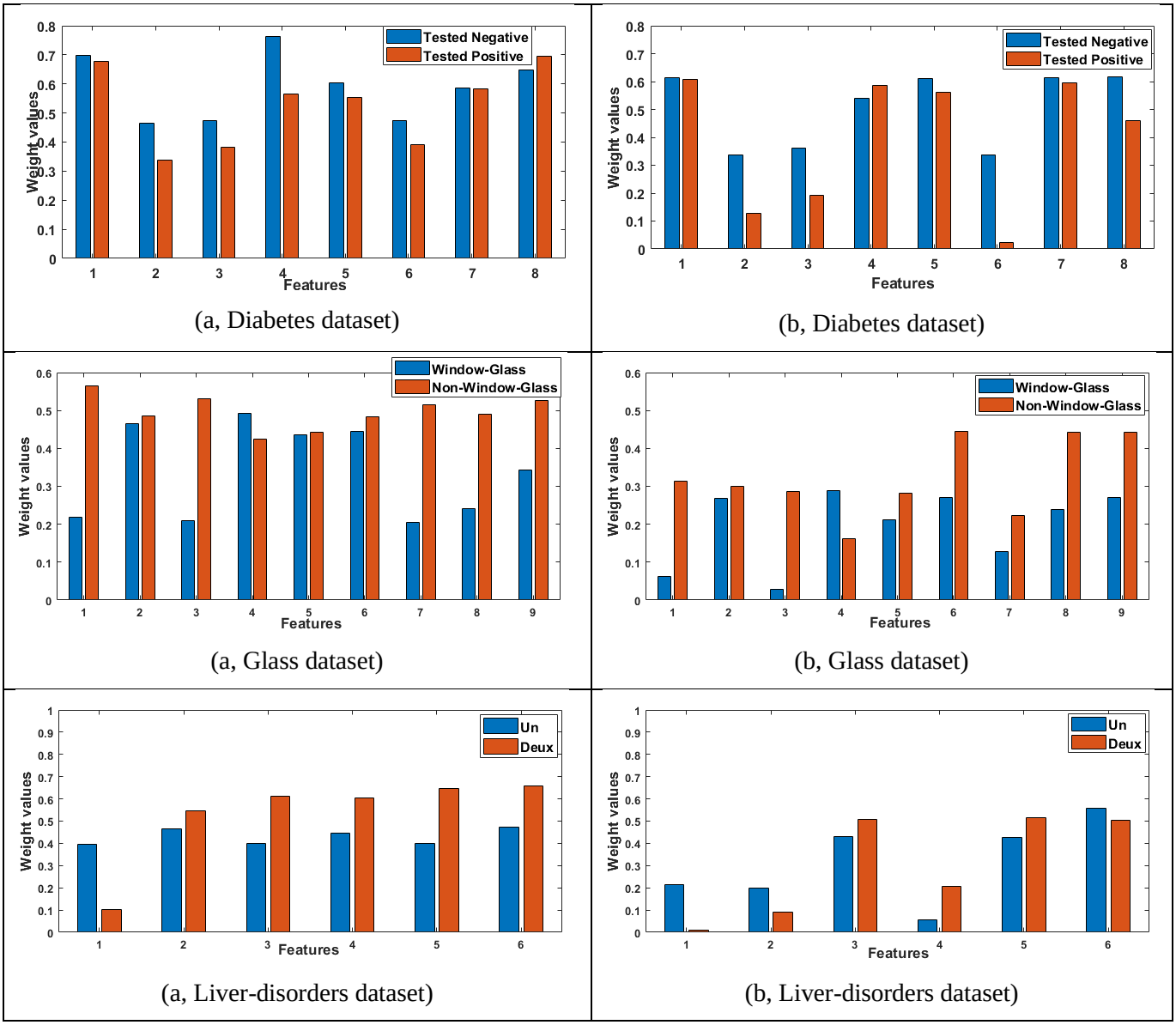
<u>Step 3:</u> Computation of the weights for distributions

Weights have been estimated based on credibility and separability degrees. That explains the non-linearity in Figure 15. By looking at Figure 12, on the information quantity ($NegEnt$), and Figure 15, on the weight values, we can see that the weight computation is more sensitive to the variation of the information quantity in the possibilistic framework than the probabilistic framework. For instance, in the probabilistic modeling of the Iris dataset, $NegEnt$ generated by the feature 2 for class 3 is larger than the one obtained by the same feature in class 2. The weight given for this feature in class 3 is smaller than the one given for the same feature in class 2. In the possibilistic modeling, the $NegEnt$ generated by the same feature 2 for class 3 is larger than the one obtained by the same feature in class 2. The weight given for this feature in class 3 is larger than the one given for the same feature in class 2.

In the probabilistic modeling of the Diabetes dataset, $NegEnt$ generated by feature 8 for the first class is larger than $NegEnt$ for the second class. Contrariwise, the weight given for feature 8 in the first class is greatly smaller than the one given for the second class. In the possibilistic modeling, for the same feature 8 of the same Diabetes dataset, the value $NegEnt$ generated by feature 8 for the first class is greatly larger than the $NegEnt$ for the second class. Also, the weight given for feature 8 in the first class is greatly larger than the one given for the second class.



(a, Iris dataset) (b, Iris dataset)



(a, Diabetes dataset) (b, Diabetes dataset)

(a, Glass dataset) (b, Glass dataset)

(a, Liver-disorders dataset) (b, Liver-disorders dataset)

Figure 15. Weight degrees in (a) probabilistic framework and (b) possibilistic framework

<u>Step 4:</u> Definition of all subsets of the distribution set Z_i

The Iris dataset contains 4 features, in this case 15 subsets are possible ($B_{k=1\ to\ 15}$). The Pima Indians Diabetes dataset contains 8 features, in this case 255 subsets are possible ($B_{k=1\ to\ 255}$) . The Glass dataset contains 9 features, in this case 512 subsets are possible ($B_{k=1\ to\ 512}$). The Liver-disorders dataset contains 6 features, in this case 64 subsets are possible ($B_{k=1\ to\ 64}$).

Step 5: Initial fusion of the weighted distributions of each subset

For each class C_i, once the weight of each feature and the different subsets $B_{i,k}$ are defined, we proceed to the fusion of the weighted distributions of each subset. Then, the obtained distributions are used to select the best subset (s), for each class C_i.

In next steps (Step6, step 7 and step8), we present by figures (Figure 16, Figure 17 and Figure 18) only results obtained for Iris dataset because it contains a low number of subsets which facilitate results visualisation and interpretation.

Step 6: Computation of the information quantity ($NegEnt$) in each subset ($B_{i,k,\ k=1\ to\ 2^m,\ m=number\ of\ features}$)

The information quantity ($NegEnt$) obtained for each subset in the possibilistic and probabilistic frameworks are presented in Figure 16. In addition, the difference between the largest value of $NegEnt$ and the smallest one is very small in the probabilistic framework unlike the possibilistic framework where the difference is more important. The value of $NegEnt$ for class 3 (yellow curve) is always the weakest one. Furthermore, the behavior of $NegEnt$ for the first class (blue curve) is always the strongest one. The behavior of $NegEnt$ for the first class and the second class (blue and red curves) are very close. It means that the construction of subsets greatly influences the separability between the second and the third class.

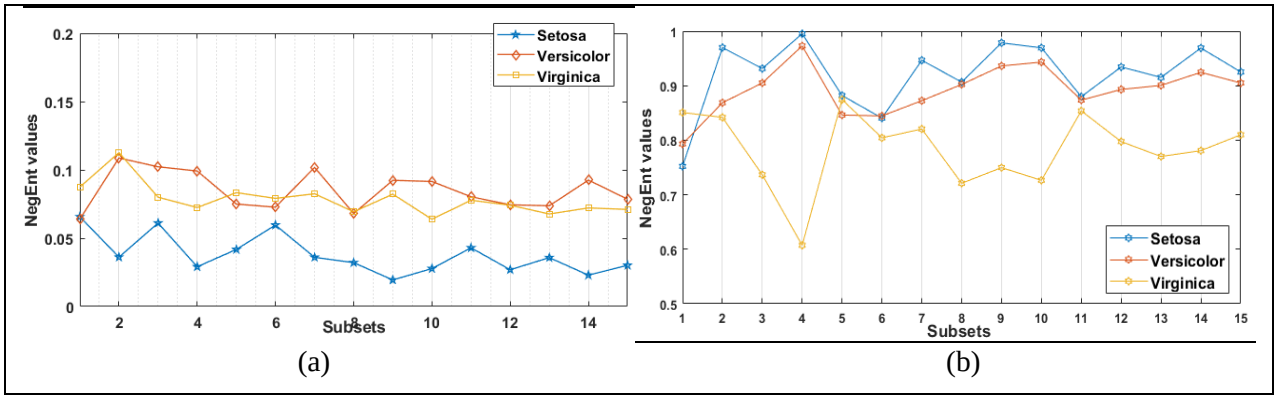


Figure 16. Information quantity ($NegEnt$) of subsets in (a) probabilistic framework and (b) possibilistic framework for Iris dataset

Step 7: Computation of the credibility measure for each subset ($B_{k=1\ to\ N}$)

The credibility measure obtained for each subset in the possibilistic and probabilistic frameworks are presented in Figure 17. It is clear that $Cred$ behavior is nonlinear when comparing both frameworks. For example, the first class $Cred$ value is always below the other classes in the probabilistic framework, but in the possibilistic framework and especially for the fourth and the ninth subsets, the first class $Cred$ value goes up to situate between the second and the third class.

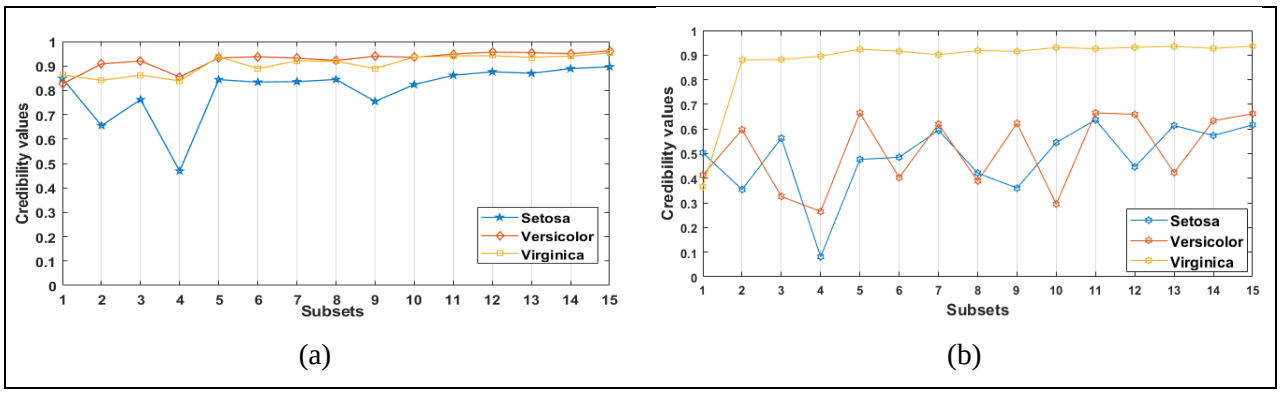


Figure 17. Credibility measures of subsets in (a) probabilistic framework and (b) possibilistic framework for Iris dataset

<u>Step 8:</u> Selection of the best subsets for the final fusion

As in Step 3, weights are estimated based on the credibility and the separability degrees, which explains the non-linearity in Figure 18. In the probabilistic framework, large weight values are applied to all subsets of the second class, and small weight values are assigned to a majority of subsets of the first class. In possibilistic framework, large weight values are applied to all subsets of the third class, and very similar weight values in each subset for the first and second classes.

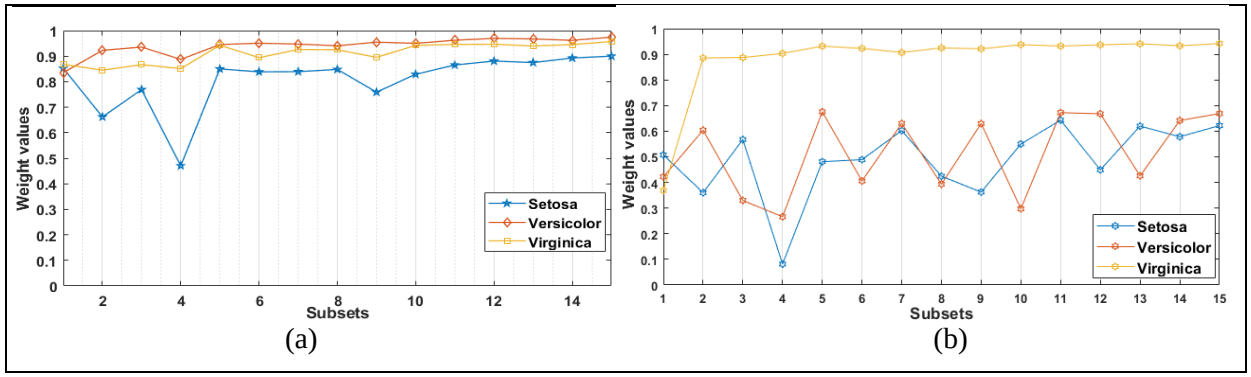


Figure18. Weight values of subsets in (a) probabilistic framework and (b) possibilistic framework for Iris dataset

Table 12, Table 13, Table 14 and Table 15 present the selected subsets respectively for the four benchmarks (Iris, Diabetes, Glass, Liver-disorders) to be considered in the final fusion step. The selection process is based on the dominance rule that uses $Cred$ and $NegEnt$ values. Table 12, Table 13, Table 14 and Table 15 show the results. The selected subsets are not the same in all classes for both frameworks.

Table 12. Selected subsets of Iris dataset

Possibilistic framework
$\mathbf{C_1}$: **Setosa**
$ND = \{B_7, B_9, B_{10}, B_{11}, B_{14}, B_{15}\}$
$\mathbf{C_2}$: **Versicol**
$ND = \{B_9, B_{11}, B_{14}, B_{15}\}$
$\mathbf{C_3}$: **Virginica**
$ND = \{B_5, B_{11}, B_{15}\}$
Probabilistic framework
$\mathbf{C_1}$: **Setosa**
$ND = \{B_1, B_5, B_{12}, B_{15}\}$
$\mathbf{C_2}$: **Versicol**
$ND = \{B_3, B_6, B_{11}, B_{15}\}$
$\mathbf{C_3}$: **Virginica**
$ND = \{B_3, B_4, B_6, B_{10}, B_{13}, B_{15}\}$

Table 13. Selected subsets of Diabetes dataset

Possibilistic framework
$\mathbf{C_1}$: **Tested Negative**
$ND = \left\{ \begin{matrix} B_5, B_7, B_8, B_{28}, B_{34}, B_{36}, B_{53}, B_{63}, B_{64}, B_{79}, B_{88}, B_{91}, B_{102}, B_{113}, B_{124}, B_{138}, B_{139}, B_{144}, B_{154}, \\ B_{155}, B_{156}, B_{168}, B_{177}, B_{178}, B_{179}, B_{192}, B_{193}, B_{194}, B_{216}, B_{219}, B_{221}, B_{225}, B_{231}, B_{240}, B_{241}, \\ B_{242}, B_{245}, B_{247}, B_{248}, B_{251}, B_{254}, B_{255} \end{matrix} \right\}$
$\mathbf{C_2}$: **Tested Positive**
$ND = \left\{ \begin{matrix} B_5, B_6, B_{23}, B_{29}, B_{64}, B_{71}, B_{116}, B_{117}, B_{122}, B_{125}, B_{165}, B_{169}, B_{173}, B_{176}, B_{182}, B_{183}, B_{205}, \\ B_{211}, B_{219}, B_{220}, B_{221}, B_{222}, B_{225}, B_{228}, B_{231}, B_{236}, B_{240}, B_{243}, B_{247}, B_{254}, B_{255} \end{matrix} \right\}$
Probabilistic framework
$\mathbf{C_1}$: **Tested Negative**
$ND = \left\{ \begin{matrix} B_5, B_8, B_{12}, B_{15}, B_{18}, B_{33}, B_{34}, B_{46}, B_{78}, B_{81}, B_{96}, B_{130}, \\ B_{135}, B_{137}, B_{187}, B_{201}, B_{224}, B_{226}, B_{230}, B_{239}, B_{255} \end{matrix} \right\}$
$\mathbf{C_2}$: **Tested Positive**
$ND = \{B_1, B_5, B_{15}, B_{28}, B_{30}, B_{78}, B_{133}, B_{144}, B_{196}, B_{215}, B_{230}, B_{242}, B_{244}, B_{253}, B_{255}\}$

Table 14. Selected subsets of Glass dataset

Possibilistic framework
$\mathbf{C_1}$: **Window − Glass**
$ND = \{B_3, B_{11}, B_{67}, B_{150}, B_{272}, B_{282}, B_{292}, B_{297}, B_{336}, B_{398}, B_{400}, B_{405}, B_{419}, B_{511}\}$
$\mathbf{C_2}$: **Non − Window − Glass**
$ND = \{B_{219}, B_{443}, B_{484}, B_{496}, B_{500}, B_{511}\}$
Probabilistic framework
$\mathbf{C_1}$: **Window − Glass**
$ND = \begin{Bmatrix} B_2, B_6, B_8, B_{14}, B_{21}, B_{23}, B_{24}, B_{43}, B_{60}, B_{63}, B_{65}, B_{106}, B_{113}, B_{115}, \\ B_{116}, B_{143}, B_{214}, B_{217}, B_{224}, B_{484}, B_{486}, B_{500}, B_{508}, B_{510}, B_{511} \end{Bmatrix}$
$\mathbf{C_2}$: **Non − Window − Glass**
$ND = \begin{Bmatrix} B_8, B_9, B_{31}, B_{36}, B_{37}, B_{43}, B_{45}, B_{93}, B_{122}, B_{123}, B_{125}, B_{129}, B_{211}, B_{230}, B_{236}, \\ B_{245}, B_{247}, B_{250}, B_{255}, B_{302}, B_{338}, B_{367}, B_{368}, B_{377}, B_{426}, B_{446}, B_{451}, B_{456}, \\ B_{457}, B_{492}, B_{496}, B_{507}, B_{511} \end{Bmatrix}$

Table 15. Selected subsets of Liver-disorders dataset

Possibilistic framework
$\mathbf{C_1}$: **Un**
$ND = \{B_{23}, B_{42}, B_{44}, B_{57}, B_{58}, B_{63}\}$
$\mathbf{C_2}$: **Deux**
$ND = \{B_1, B_{10}, B_{22}, B_{26}, B_{29}, B_{35}, B_{57}, B_{63}\}$
Probabilistic framework
$\mathbf{C_1}$: **Un**
$ND = \{B_5, B_6, B_{15}, B_{16}, B_{19}, B_{20}, B_{31}, B_{34}, B_{37}, B_{40}, B_{41}, B_{56}, B_{62}, B_{63}\}$
$\mathbf{C_2}$: **Deux**
$ND = \{B_3, B_5, B_{15}, B_{28}, B_{31}, B_{37}, B_{40}, B_{46}, B_{62}, B_{63}\}$

To verify the quality of a subset, a score associated with a result of fusion is calculated as in Section 7.1. The score associated with each subset is illustrated in Figure 19. Selected subsets according to the dominance rule are circled in red. Subsets with the highest scores are selected by the dominance rule irrespective of the framework.

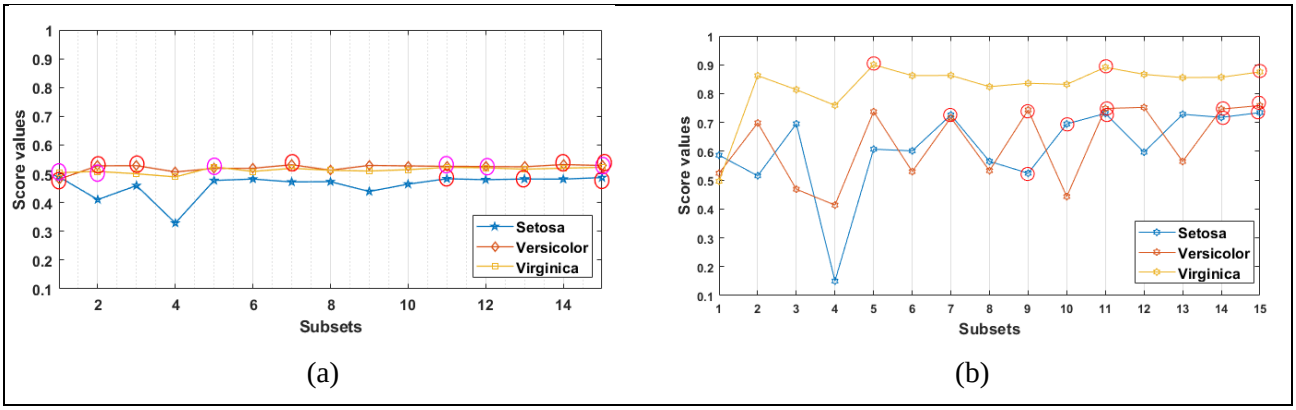


Figure 19. Scores of subsets in (a) probabilistic framework and (b) possibilistic framework for Iris dataset

Finally, the recognition rate is calculated based on SVM (Support Vector Method) classifier using 10-fold cross-validation based on all features, when the SVM classifier is applied without subset selection, or the union of the selected subsets for each class whether in the probabilistic or possibilistic framework. The use of the possibility theory has shown its advantages in data modeling compared to the probability theory and to the SVM classifier as illustrated in Table 16. This advantage is quite important for the IRIS database, this is mainly due to the reduced size of this database. In fact, as we have previously noted, the IRIS database consists of only 150 samples that are equally divided between the different classes. Then, the samples from each class are further divided into two parts, the first one for training and the other for testing. We can then conclude that the number of samples is insufficient to build and learn a probabilistic model. On the other hand, possibilistic theory and SVM classifier have been able to cope with that poor-data small sample size situation to build a model capable of generalizing the classification problem of IRIS.

The Diabetes database, Glass database and Liver-disorders database contain more samples than the IRIS database. The number of samples seems enough to construct a probabilistic model that gives similar recognition rates that with possibilistic modeling and with SVM classifier (79.3% versus 80.6% and 78.1% for Diabetes database, 94.5% versus 97.3% and 95.3% for Glass database and 66.6% versus 69.9% and 67.3% for Liver-disorders database). The possibility theory is an uncertainty theory that deal better with incomplete information [33] than the probability theory and the classical SVM classifier. This is even more confirmed in a scenario where we deliberately impoverish the Diabetes dataset, the Glass dataset and Liver-disorders dataset. The poorer Diabetes dataset version contains only 50 samples for each class. The poorer Glass dataset and Liver-disorders dataset versions contain only 20 samples for each class. The recognition rates, illustrated in Table 16 for the impoverished datasets, show the advantage of the possibility theory in data modeling compared to the probability theory and the SVM classifier in poor data environments (information incompleteness). Same kind of advantages have been confirmed from results that have been obtained for different kind of applications namely in pattern recognition and image segmentation in the processing of poor-quality images [32, 34-37].

It is worth mentioning that the size of databases in Table 16 does not exceed 800 samples. We notice from Table 16 that for some classes, there is no big difference between the number of samples used before and after the database impoverishment scheme. As an example, the number of samples from Class 2 of the Diabetes database, initially 50, is reduced to 30 after the impoverishment process. Similarly, in the case of Class 2 of the Glass database, the number is 15 samples before the reduction process and 12 samples after. We also observe an imbalance between the number of samples for the classes within the same database. To overcome these inherent difficulties to classification of real databases, we propose to make a performance comparison with a very large database where the classes are balanced with respect to number of samples.

Table 16. Performance comparison of the possibilistic modelling with the probabilistic one and SVM classifier

	Iris dataset	Diabetes dataset	Impoverished Diabetes dataset	Glass dataset	Impoverished Glass dataset	Liver-disorders dataset	Impoverished Liver-disorders dataset
Feature number	4	8	8	9	9	6	6
Class number	3	2	2	2	2	2	2
Total sample size / Class	50	500 /C1 268 /C2	50	163 /C1 51 /C2	20	138 /C1 188 /C2	20
Training sample size / Class	35	100 /C1 50 /C2	30	58 /C1 15 /C2	12		12
Possibilistic	**96%**	**80.6%**	**73%**	**97.3%**	**100%**	**69.9%**	**71.5%**
Probabilistic	80%	79.3%	63%	94.5%	92.8%	66.6%	64%
SVM Classifier	**96%**	78.1%	69%	95.3%	92.5%	67.3%	69%

8.2.3 Comparison of Probabilistic-Possibilistic methodology on large sample size

In this section, a large database, namely Skin dataset, is considered. The Skin database has two classes: the Skin class and the non-skin class, both described by three attributes [38] [39]. The total sample size of this base is 245057, of which 50859 are skin samples and 194198 are non-skin samples. The imbalance between the two classes, Skin and non-Skin, is also present in this database. The database is balanced by taking 50 000 samples for each class. The impoverishment process is implemented as follows: starting with 40 000 samples per class down to 5 samples per class. The sizes considered are then: 50 000, 20 000, 10 000, 5000, 2500, 20, 15, 10 and 5. For each size, a Monte Carlo process (1000 draws) has been applied to obtain results for the three approaches: the Yager & Petry [1] probabilistic approach, its corresponding possibilistic approach (as proposed in this paper) and a classical SVM classifier.

The following steps have been performed to compare the three approaches:

- Select data subsets, by the possibilistic approach taking into account the quality of the attributes, and then applying the SVM approach in the classification stage;
- Select data subsets, using the probabilistic approach taking into account the quality of the attributes, and then applying the SVM approach in the classification stage.
- Use of the three attributes of the database, without subset selection and not taking into account of attribute quality, and then applying the SVM approach in the classification stage.

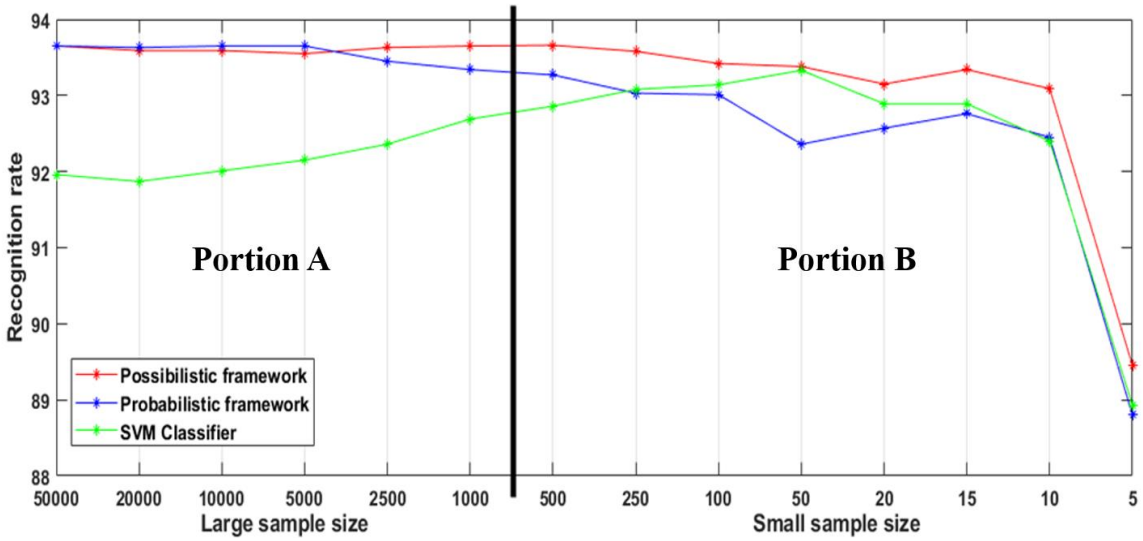


Figure 20. Performance comparison of the possibilistic framework with the probabilistic one and SVM classifier using the large size Skin dataset

The classification rate is calculated using the principle of cross validation (10 times). An average classification rate for 1000 draws of each size has been calculated. Figure 20 illustrates the performance achieved by the three approaches for each size. The figure is presented in two parts (Portion A, Portion B) where "Portion A" is for large size samples (50000, 20000, 10000, 5000, 2500, 1000) and "Portion B" is for small size samples (500, 250, 100, 50, 20, 15, 10 and 5).

In "Portion A" of figure 20, the probabilistic approach gives excellent results in terms of classification rates as well as for the possibilistic approach. The performance obtained with the classical SVM is less good as compared to the two other ones in "Portion A" but it is known that SVM is better designed for small size cases as exhibited in "Portion B". In "Portion B", the small size cases, the possibilistic approach outperforms gradually the other two methods as the data impoverishment augments. This is under the condition where the possibilistic method can be exploited at best. Finally, it is worth mentioning that the possibilistic method demonstrates the best potential of the three in a context where we would need to reduce the size of the data while maintaining excellent performances.

9. Conclusion

This paper offers a companion paper to Yager and Petry who proposed a quality-based methodology to combine data provided by multiple probabilistic sources to improve quality of information for decision-makers. In this paper, we adapted the Yager-Petry 8-steps methodology, not identified as such in Yager and Petry, to be used with multiple possibilistic sources. The paper offers a side by side development of the possibilistic approach with respect to the probabilistic one where both approaches are compared by the means of four experimental benchmark datasets: one being data-poorer (IRIS data set) than the others one (Diabetes dataset, Glass dataset and Liver-disorders dataset). Results obtained from comparison of both approaches confirm the superiority of the possibilistic approach in the presence of less complete information. Considering the results of this paper, incompleteness of information is the only aspect that we can objectively bring concluding remarks upon. Other aspects will obviously have to be taken into account in the choice of a modelling approach in the presence of a diversity of information sources (hard and soft) such as the nature and characteristics of the sources as well as the specifics of the domain of applications.

References

[1] R. R. Yager and F. Petry, "An intelligent quality-based approach to fusing multi-source probabilistic information," *Information Fusion,* vol. 31, pp. 127-136, 2016.

[2] F. Herrera, S. Alonso, F. Chiclana, and E. Herrera-Viedma, "Computing with words in decision making: foundations, trends and prospects," *Fuzzy Optimization and Decision Making,* vol. 8, pp. 337-364, 2009.

[3] E. P. Blasch, D. A. Lambert, P. Valin, M. M. Kokar, J. Llinas, S. Das*, et al.*, "High Level Information Fusion (HLIF): Survey of models, issues, and grand challenges," *Aerospace and Electronic Systems Magazine, IEEE,* vol. 27, pp. 4-20, 2012.

[4] D. L. Hall and J. M. Jordan, *Human-Centered Information Fusion*: Artech House, Incorporated, 2010.

[5] P. Walley, *Statistical Reasoning With Imprecise Probabilities*: Chapman and Hall, 1991.

[6] G. Shafer, *A Mathematical Theory of Evidence*: Princeton University Press, 1976.

[7] L. Zadeh, "Fuzzy Sets as the Basis for a Theory of Possibility," *Fuzzy Sets and Systems,* vol. 1, pp. 3-28, 1978.

[8] R. R. Yager, "Hard and soft information fusion using measures," in *2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering*, 2010, pp. 13-16.

[9] M. Liggins, D. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice, Second Edition*: Taylor & Francis, 2008.

[10] D. L. Hall and S. A. H. McMullen, *Mathematical Techniques in Multisensor Data Fusion*: Artech House, 2004.

[11] B. Khaleghi, A. Khamisa, F. O. Karraya, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion,* vol. 14, pp. 28-44, 2013.

[12] S. Das, *High-Level Data Fusion*: Artech House, 2008.

[13] E. Blasch, É. Bossé, and D. A. Lambert, Eds., *High-Level Information Fusion Management and Systems Design*. Artech House, 2012, p.^pp. Pages.

[14] D. Dubois, W. Liu, J. Ma, and H. Prade, "The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks," *Information Fusion,* vol. 32, pp. 12-39, 2016.

[15] L. Snidaro, J. Garcia-Herrera, J. Llinas, and E. Blasch, "Context-Enhanced Information Fusion," in *Boosting Real-World Performance with Domain Knowledge*, ed: Springer, 2016.

[16] J.-M. Tacnet, S. Carladous, J. Dezert, and M. Batton-Hubert, "New integrated and multiscale decision-aiding framework in a context of imperfect information: application to the assessment of torrent checkdams' effectiveness," in *EGU General Assembly Conference Abstracts*, 2017, p. 4440.

[17] J.-M. Tacnet, M. Batton-Hubert, and J. Dezert, "A two-step fusion process for multi-criteria decision applied to natural hazards in mountains," *arXiv preprint arXiv:1005.0896,* 2010.

[18] É. Bossé and G. Rogova, Eds., *Information Quality in Information Fusion and Decision Making* (Information Fusion and Data Science. Switzerland: Springer Nature, 2019, p.^pp. Pages.

[19] S. Calderwood, K. McAreavey, W. Liu, and J. Hong, "Context-dependent combination of sensor information in Dempster–Shafer theory for BDI," *Knowledge and Information Systems,* vol. 51, pp. 259-285, 2017.

[20] G. J. Klir and J. F. Geer, "Information-Preserving Probability-Possibility Transformations," in *Fuzzy logic*, ed: Springer, 1993, pp. 417-428.

[21] D. Dubois and E. Hüllermeier, "A notion of comparative probabilistic entropy based on the possibilistic specificity ordering," in *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, 2005, pp. 848-859.

[22] P. Elmore and F. Petry, "Information Theory Applications in Soft Computing," in *Granular, Soft and Fuzzy Approaches for Intelligent Systems*, ed: Springer, 2017, pp. 81-97.

[23] R. R. Yager, "On prioritized multiple-criteria aggregation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 42, pp. 1297-1305, 2012.

[24] R. A. Fisher. *IRIS data set*. Available: https://archive.ics.uci.edu/ml/datasets/Iris

[25] S. Lekkas and L. Mikhailov, "Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases," *Artificial Intelligence in Medicine,* vol. 50, pp. 117-126, 2010.

[26] P. I. D. dataset. (5 March 2019). *Pima Indians Diabetes dataset*. Available: http://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetes.data

[27] K. Bache and M. Lichman. (2013, 28 April 2019). *Glass Data Set* Available: https://archive.ics.uci.edu/ml/machine-learning-databases/glass/

[28] K. Bache and M. Lichman. (2013, 28 April 2019). *Liver-disorders data set*. Available: https://archive.ics.uci.edu/ml/machine-learning-databases/liver-disorders/

[29] D.-J. Chang, A. H. Desoky, M. Ouyang, and E. C. Rouchka, "Compute pairwise manhattan distance and pearson correlation coefficient of data points with gpu," in *2009 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing*, pp. 501-506, 2009.

[30] D. Dubois, H. Prade, and S. Sandri, "On possibility/probability transformations," in *Fuzzy logic*, ed: Springer, pp. 103-112, 1993.

[31] A. J. Pinar, D. T. Anderson, T. C. Havens, A. Zare, and T. Adeyeba, "Measures of the Shapley index for learning lower complexity fuzzy integrals," *Granular Computing,* vol. 2, pp. 303-319, 2017.

[32] S. A. Bouhamed, I. K. Kallel, D. S. Masmoudi, and B. Solaiman, "Feature selection in possibilistic modeling," *Pattern Recognition,* vol. 48, pp. 3627-3640, 2015.

[33] P. Agarwal and H. Nayal, "Possibility theory versus probability theory in fuzzy measure theory," *Int. Journal of Engineering Research and Applications,* vol. 5, pp. 37-43, 2015.

[34] I. K. Kallel, S. Almouahed, B. Solaiman, and É. Bossé, "An iterative possibilistic knowledge diffusion approach for blind medical image segmentation," *Pattern Recognition,* vol. 78, pp. 182-197, 2018.

[35] B. Alsahwa, B. Solaiman, S. Almouahed, E. Bosse, and D. Gueriot, "Iterative refinement of possibility distributions by learning for pixel-based classification," *IEEE Transactions on Image Processing,* vol. 25, pp. 3533-3545, 2016.

[36] B. Alsahwa, B. Solaiman, É. Bossé, S. Almouahed, and D. Gueriot, "A method of spatial unmixing based on possibilistic similarity in soft pattern classification," *Fuzzy information and engineering,* vol. 8, pp. 295-314, 2016.

[37] I. K. Kallel, S. Almouahed, B. Alsahwa, and B. Solaiman, "The use of contextual spatial knowledge for low-quality image segmentation," *Multimedia Tools and Applications,* pp. 1-21, 2018.

[38] R. B. Bhatt, G. Sharma, A. Dhall, S. Chaudhury, "Efficient skin region segmentation using low complexity fuzzy decision tree model", *IEEE-INDICON* 2009, Dec 16-18, Ahmedabad, India, pp.1-4, 2009.

[39] A. Dhall, G. Sharma, R. Bhatt, G. M. Khan, "Adaptive Digital Makeup", *in Proc. of International Symposium on Visual Computing (ISVC)* 2009, Nov. 30 Dec. 02, Las Vegas, Nevada, USA, Lecture Notes in Computer Science, Vol. 5876, pp. 728-736, 2009.