IEEE Access
SPECIAL SECTION ON MACHINE LEARNING DESIGNS, IMPLEMENTATIONS AND TECHNIQUES

Received February 3, 2020, accepted March 2, 2020, date of publication March 9, 2020, date of current version March 19, 2020.
Digital Object Identifier 10.1109/ACCESS.2020.2979553

INVITED PAPER

Possibilistic Similarity Measures for Data Science and Machine Learning Applications

AMAL CHARFI[1], SONDA AMMAR BOUHAMED[1,2], ÉLOI BOSSÉ[2,3], (Senior Member, IEEE), IMENE KHANFIR KALLEL[1,2], WASSIM BOUCHAALA[4], BASEL SOLAIMAN[2], AND NABIL DERBEL[1]
[1]Control and Energy Managment (CEM Lab), National School of Engineers of Sfax, University of Sfax, Sfax 3038, Tunisia
[2]Image & Information Processing Department (iTi), IMT Atlantique, 83818-29238 Brest, France
[3]Expertises Parafuse, Inc., Quebec, QC G1W 4N1, Canada
[4]Tunisian Professional Training Agency, Sfax 3000, Tunisia

Corresponding author: Éloi Bossé (ebosse861@gmail.com)

ABSTRACT Measuring similarity is of a great interest in many research areas such as in data sciences, machine learning, pattern recognition, text analysis and information retrieval to name a few. Literature has shown that possibility is an attractive notion in the context of distinguishability assessment and can lead to very efficient and computationally inexpensive learning schemes. This paper focuses on determining the similarity between two possibility distributions. A review of existing similarity measures within the possibilistic framework is presented first. Then, similarity measures are analyzed with respect to their capacity to satisfy a set of required properties that a similarity measure should own. Most of the existing possibilistic similarity measures produce undesirable outcomes since they generally depend on the application context. A new similarity measure, called InfoSpecificity, is introduced and the similarity measures are categorized into three main methods: morphic-based, amorphic-based and hybrid. Two experiments are being conducted using four benchmark databases. The aim of the experiments is to compare the efficiency of the possibilistic similarity measures when applied to real data. Empirical experiments have shown good results for the hybrid methods, particularly with the InfoSpecificity measure. In general, the hybrid methods outperform the other two categories when evaluated on small-size samples, i.e., poor-data context (or poor-informed environment) where possibility theory can be used at the greatest benefit.

INDEX TERMS Classification, distance, entropy, learning, measures of specificity, possibility distributions, similarity, uncertainty.

I. INTRODUCTION

Determining similarities is part of a fundamental process of a human sense-making mechanism that consists of three elements: an object or event, a mental model, and an association between them [1]. Similarities represent the core of that association scheme. The notion of similarity has been exploited in various fields of Computer Sciences [2]–[5] such as in machine learning pattern recognition [4], classification [6], image processing [7] and decision making [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Lara.

Similarity in a machine learning context is required to compute the "closeness" between elements in a dataset. It allows to understand the structure within the input data [8]. Moreover, it is widely used in classification tasks in order to find the best match between new instances and the already known ones (by training) [9]. Several researchers in the machine learning field have an ongoing interest on similarity methods [8]–[12]. Refining the estimation of similarity scores leads to the improvement of algorithms accuracy as well as the minimization of errors and confusions.

The similarity notion is useful to describe the degree of relationship between concepts or entity models. It represents a crucial tool to ensure comparison, matching, discrimination, sets distinguishability, determination of equivalence, and classification. Therefore, a measure of similarity must be sensitive to the nature of the data and must have the ability to describe existing links between data sets. Significant effort [2], [13]–[15] has been done to develop adequate measures tailored to specific problems.

The definition of standard similarity measures has been conducted in situations where data is accurate, unambiguous and complete. However, in real-world applications, we are dealing with imprecise, uncertain or even missing data. This can be linked to a natural variability in an experiment measurement (random uncertainty) or to the lack of knowledge of the modeled phenomenon (epistemic uncertainty). Thus, these types of imperfections should be taken into account in the definition of a measure within an appropriate framework that best models these imperfections.

Probability theory is the most traditional tool for modeling imperfect information. Its major concern is to represent the variability of a phenomenon. Other types of imperfections such as the epistemic uncertainties or the incompleteness of a data set is hardly modeled by this theory. This has led to the development of new uncertainty theories like P-boxes [16], fuzzy set theory [17], evidence theory [18] and possibility theory [19].

In this paper, we are interested in studying similarity measures in a context where the information is characterized by epistemic uncertainty. Possibility theory offers an adequate representation of this kind of uncertainty. Literature has shown that possibility is an attractive notion within the context of distinguishability assessment and leads to very efficient and computationally inexpensive learning schemes [20]. The study starts by comparing existing similarity measures defined within the possibility theory framework. The existing measures have been categorized into amorphic, morphic and hybrid ones depending on what aspects they are addressing with respect to similarity: amplitude, shape, color, functions, etc.

The paper is organized as follows. Section 2 briefly presents the possibility theory. Section 3, first, presents the existing possibilistic similarity measures and, then, proposes a new measure called InfoSpecificity. Section 4 shows the results of a numerical analysis comparison of morphic and magnitude-based (amorphic) similarity measures. Section 5 presents the performance of the similarity measures when tested against real data.

II. PRELIMINARIES ON POSSIBILITY THEORY

Possibility theory is derived from the fuzzy set theory [19]. It was defined by Zadeh in 1978 and further developed by Dubois and Prade [21] since 1988. The possibilistic framework can deal with both imprecision and incomplete information through a possibility distribution π and two set measures: a possibility measure Π, and a necessity measure N.

Let Ω be a universe of discourse (or possible states) composed of n elements $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$.

A. POSSIBILITY DISTRIBUTION

The possibility distribution, denoted π, is a function $\pi : \Omega \mapsto [0, 1]\,; \omega_i \mapsto \pi(\omega_i)$ that associates a possibility degree to each element ω_i of the universe Ω of possible states.

B. POSSIBILITY AND NECESSITY MEASURES

The measure of possibility Π of an event A is a function which associates to A, a coefficient (i.e.,possibility degree) between 0 and 1:

$$\Pi(A) = \{\sup[\pi(x), x \in A]\} \tag{1}$$

The necessity measure is defined as the measure of the impossibility of the opposite event:

$$\forall A \subseteq \Omega, N(A) = 1 - \Pi\left(\overline{A}\right) \tag{2}$$

where $\Pi\left(\overline{A}\right)$ is the possibility of the opposite event.

C. NORMALIZATION

A normalized possibility distribution is assumed to contain at least one element $\omega_i \in \Omega$ which is fully possible, i.e., $\pi(\omega_i) = 1$. If not, the possibility distribution is considered as being inconsistent. The degree of inconsistency of a possibility distribution denoted $\mathrm{Inc}(\pi)$ is defined as being the maximal possibility degree:

$$\mathrm{Inc}(\pi) = \max_{\omega_i \in \Omega}\{\pi(\omega_i)\} \tag{3}$$

D. PRINCIPLE OF MINIMUM OF SPECIFICITY

Possibility theory is driven by the principle of minimal specificity [22]. In fact, a distribution π_1 is said to be more specific than a distribution π_2 if and only if $\forall \omega_i \in \Omega, \pi_1(\omega_i) \leq \pi_2(\omega_i)$. This implies that the most specific distribution is the most informative one.

E. UNCERTAINTY MEASURES

To evaluate the amount of information in a possibility distribution, several uncertainty measures have been proposed in the literature [23]. In this section, we limit the discussion to the following measures: the U-Uncertainty measure [24] and the Specificity measure [22] and we consider $\pi = \{\pi(\omega_1), \pi(\omega_2), \ldots, \pi(\omega_n)\}$ as an ordered possibility distribution such as $\pi(\omega_1) \geq \pi(\omega_2) \geq \ldots \geq \pi(\omega_n)$.

1) U-UNCERTAINTY MEASURE

The U-uncertainty measure [18], [19] is given by the following equation:

$$U(\pi) = \left[\sum_{i=1}^{n} (\pi(\omega_i) - \pi(\omega_{i+1})) \log_2(i)\right] + \left[(1 - \pi(\omega_1)) \log_2(n)\right] \tag{4}$$

with $\pi(\omega_{n+1}) = 0$.

Note that the range of U is $\left[0, \log_2(n)\right]$, with $U(\pi) = 0$ in the case of "complete knowledge" and $U(\pi) = \log_2(n)$ in the case of "complete ignorance". Also, the term $(1 - \pi(\omega_1)) \log_2(n)$, provides a generalization for subnormalized distributions.

2) SPECIFICITY MEASURE

The specificity measure [22] is a function defined by Sp : $\pi(\Omega) \in [0, 1]$ where:

$$\mathrm{Sp}(\pi) = \pi(\omega_1) - \sum_{j=2}^{n} p_j \pi(\omega_j) \quad (5)$$

with p_j is a weight such that $p_j \in [0, 1]$. In addition, it must satisfy: $p_j \geq p_i, \forall\ 1 < j < i$ and $\sum_{j=2}^{n} p_j = 1$. In fact, the specificity measure of a possibility distribution decreases when the degree of possibility of its elements is increasing: if $\pi(\omega_1) \geq \pi(\omega_2)$, then $\mathrm{Sp}(\pi(\omega_1)) \leq \mathrm{Sp}(\pi(\omega_2))$. Finally, the specificity measure presents the following two extreme cases:

- Complete knowledge: $\mathrm{Sp}(\pi) = 1$
- Complete ignorance: $\mathrm{Sp}(\pi) = 0$

III. EXISTING POSSIBILISTIC SIMILARITY MEASURES AND A NEW ONE: InfoSpecificity

The comparison of data sets of uncertain information depends on their representation models. In particular, comparing uncertain information in the possibility framework reverts to compare their possibility distributions. Therefore, a similarity measure must be able to evaluate the similarity between two possibility distributions.

A. BASIC MATHEMATICAL PROPERTIES OF A POSSIBILISTIC SIMILARITY MEASURE

A similarity measure between two normalized possibility distributions is a function $s : \pi(\Omega)^2 \mapsto [0, 1]$;

$(\pi_i, \pi_j) \mapsto s(\pi_i, \pi_j)$ that must satisfy a set of basic properties as listed below [10], [26]. Let π_1 and π_2, be two possibility distributions.

Property 1: Non-negativity

$$s(\pi_1, \pi_2) \geq 0 \quad (6)$$

Property 2: symmetry

$$s(\pi_1, \pi_2) = s(\pi_2, \pi_1) \quad (7)$$

Property 3: Upper bound

$$\forall \pi_i, s(\pi_i, \pi_i) = 1, \text{ and } \forall \pi_i, \pi_j, s(\pi_i, \pi_j) \leq 1 \quad (8)$$

The upper bound is equal to 1 and this bound is obtained only if all the elements of the two distributions are identical.

Property 4: Lower Bound

$$\begin{aligned} &\text{If } \forall \omega_i \in \Omega, \pi_1(\omega_i) \in \{0, 1\}, \pi_2(\omega_i) \in \{0, 1\}, \\ &\quad \pi_2(\omega_i) = 1 - \pi_1(\omega_i), \\ &\qquad \text{then } s(\pi_1, \pi_2) = 0 \end{aligned} \quad (9)$$

The lower bound is equal to 0 and this bound is reached only when the compared distributions are contradictory.

Property 5: Specificity

The similarity measure should satisfy the principle of minimum of specificity:

$$\begin{aligned} &\text{If } \forall \omega_i \in \Omega, \pi_1(\omega_i) \leq \pi_2(\omega_i), \pi_2(\omega_i) \leq \pi_3(\omega_i) \\ &\qquad \text{then, } s(\pi_1, \pi_2) \geq s(\pi_1, \pi_3) \end{aligned} \quad (10)$$

Property 6: Permutation

Let π_1 and π_2 be two possible distributions and let ρ be a permutation of their indexes, then:

$$\forall \pi_1, \pi_2, s(\pi_1, \pi_2) = s(\pi_{\rho(1)}, \pi_{\rho(2)}) \quad (11)$$

So, the similarity between two distributions of possibilities π_1 and π_2 does not change if the order of the elements is changed.

B. EXISTING POSSIBILISTIC SIMILARITY MEASURES

Although, several recent works have been done on the applications of similarity measures [27]–[32] in different fields of research, relatively few ones are dedicated to measure the similarity between possibility distributions.

1) INFORMATION CLOSENESS: InfoCloseness

Higashi and Klir proposed a kind of seminal work [25] on measuring similarity between possibility distributions. This measure is based on the U-uncertainty measure. It computes the variation of information between the compared distributions. The G-similarity between two possibility distributions π_1 and π_2, denoted G, is defined as:

$$G(\pi_1, \pi_2) = g(\pi_1, \pi_1 \vee \pi_2) + g(\pi_2, \pi_1 \vee \pi_2) \quad (12)$$

with $g(\pi_j, \pi_i) = U(\pi_j) - U(\pi_i)$, $\vee$ represents the maximum operator and U is the uncertainty measure given by equation (4). Then, the similarity measure, InfoCloseness, derived from the function $G(\pi_1, \pi_2)$ is defined as follows:

$$\mathrm{InfoCloseness} = 1 - \frac{G(\pi_1, \pi_2)}{G_{\max}} \quad (13)$$

with $G_{\max} = 2 * \log_2(n) - \log_2(n-1)$ which is achieved when comparing a possibility distribution of "complete knowledge" and its complementary possibility distribution.

2) SANGÜESA DISTANCE

Sangüesa [26] has conducted a study in the context of learning possibilistic causal networks. He has defined a measure, called Sang, that represents the uncertainty of the difference between two possibility distributions.

$$\mathrm{Sang}(\pi_1, \pi_2) = U(\pi_d) \quad (14)$$

with $\pi_d(\omega) = |\pi_1(\omega_i) - \pi_2(\omega_i)|, \forall \omega_i \in \Omega$.

In order to have values in interval [0, 1], the measure, Sang, is normalized as:

$$\mathrm{Sang}(\pi_1, \pi_2) = \frac{\mathrm{Sang}(\pi_1, \pi_2)}{\max(U(\pi_d(\omega_i)))} \quad (15)$$

3) INFORMATION DIVERGENCE

The information divergence measure [33] is proposed from an analogy to the probabilistic measure of divergence. It uses discrete Choquet integral of the distribution difference π_d such that $\pi_d(\omega) = |\pi_1(\omega_i) - \pi_2(\omega_i)|, \forall \omega_i \in \Omega$.

$$\Pi_1(A_{\sigma(i+1)})] \quad (16)$$

with σ is a permutation of indexes such that: $\pi_d\left(\omega_{\sigma(i)}\right) \leq \ldots \leq \pi_d\left(\omega_{\sigma(n)}\right), A_{\sigma(i)} = \left\{\omega_{\sigma(i)}, \ldots, \omega_{\sigma(n)}\right\}, i = 1 \ldots n$ and $A_{\sigma(n+1)} = \emptyset$

$$S_D(\pi_1, \pi_2) = 1 - D(\pi_1|\pi_2) \tag{17}$$

4) DELTA δ

The similarity measure δ [34], between two distributions π_1 and π_2, is defined by:

$$\delta(\pi_1, \pi_2) = \frac{\sum_{i=1}^{n}(\pi_1(\omega_i) \wedge \pi_2(\omega_i))}{\sum_{i=1}^{n}(\pi_1(\omega_i) \vee \pi_2(\omega_i))} \tag{18}$$

where $\wedge$ represents the minimum operator and $\vee$ represents the maximum operator.

5) MINKOWSKI DISTANCE

Since possibility distributions are represented by vectors of real values in the interval [0, 1], the Minkowski distance can be applied to assess similarity in the possibility framework:

$$L_p(\pi_1, \pi_2) = 1 \sqrt[p]{\sum_{i=1}^{n} |(\pi_1(\omega_i) - \pi_2(\omega_i))|^p} \tag{19}$$

Similarity measures based on the particular cases of the Minkowski distance are:

- Similarity measure based on the normalized Manhattan distance

$$S_M(\pi_1, \pi_2) = 1 - \frac{\sum_{i=1}^{n} |(\pi_1(\omega_i) - \pi_2(\omega_i))|}{n} \tag{20}$$

- Similarity measure based on the normalized Euclidean distance;

$$S_E(\pi_1, \pi_2) = 1 - \frac{\sum_{i=1}^{n} (\pi_1(\omega_i) - \pi_2(\omega_i))^2}{n} \tag{21}$$

- Similarity measure based on the normalized Maximum distance

$$S_C(\pi_1, \pi_2) = 1 - \max_{i=1}^{n}(\pi_1(\omega_i) - \pi_2(\omega_i)) \tag{22}$$

6) INFORMATION AFFINITY: InfoAffinity

The information affinity measure [10], called here InfoAffinity, applies the concept of inconsistency to perform the degree of conflict between distributions. The proposed measure consists of combining the distance measure and the inconsistency value of the two possibility distributions:

$$\text{InfoAffinity}(\pi_1, \pi_2) = 1 - \frac{\alpha \cdot d(\pi_1, \pi_2)}{\alpha + \beta} + \frac{\beta \cdot \text{Inc}(\pi_1 \wedge \pi_2)}{\alpha + \beta}. \tag{23}$$

where $d(\pi_1, \pi_2)$ is the Manhattan distance between the two compared distributions, $\text{Inc}(\pi_1 \wedge \pi_2)$ is the inconsistency degree given by the equation (3) where $\wedge$ denotes the min operator, α and β are two coefficients such that $\alpha > 0$ and $\beta > 0$.

C. A NEW SIMILARITY MEASURE CALLED INFORMATION SPECIFICITY: InfoSpecificity

The information specificity measure, initially defined in [11], is based on two robust measures: 1) a distance measure that compares the accuracy of the information by the means of 'a point to point' scheme; and, 2) a specificity measure that allows to quantify the variation of the amount of information.

We use a standard metric distance such as the normalized Manhattan distance to compare the distributions point to point. In addition to the Manhattan distance, we exploit the information based upon the ranking of possibility degrees belonging to the distribution support. This is defined by the principle of minimum of specificity that is well illustrated by the specificity measures [35]. Then, the specificity relation between possibility distributions can be combined with the Manhattan distance, in order to qualify the native relation between the compared possibility distributions. To that end, we use the specificity measure proposed by Yager [22], given by equation (5). In fact, equation (5) offers an intuitive conception that is easy to apply. The measure belongs to the unit interval [0, 1] without a need for normalization and it satisfies the main properties of a possibilistic measure of similarity.

Nevertheless, a specificity measure, such as Yager's one, when considered alone, is not highly efficient for assessing similarity. To illustrate the issue, consider the following example.

Example: Let π_1 and π_2 be two possibility distributions such that $\pi_1 = [0, 1]$ (complete knowledge) and $\pi_2 = [1, 0]$ (complete knowledge). Using Yager's definition of a specificity measure, we obtain $\text{Sp}(\pi_1) = \text{Sp}(\pi_2) = 1$. In this case, we are facing with two different distributions and the Yager's measure gives the same amount of information.

A new similarity measure is then being defined as equation (24) below with two weights α and β that are being used to balance the contribution of: 1) a distance measure of 'a point to point' scheme; and, 2) the Yager's specificity measure quantifying the variation of the amount of information.

Definition of InfoSpecificity: Let π_1 and π_2 be two distributions defined on the same universe ω. An information specificity measure, denoted here InfoSpecificity, is proposed as follows:

$$\text{InfoSpecificity}(\pi_1, \pi_2) = \begin{cases} 1 - \dfrac{\alpha \cdot \text{dist}_{\text{Manh}}(\pi_1, \pi_2) + \beta \cdot \text{dist}_{\text{Sp}}(\pi_1, \pi_2)}{\alpha + \beta} & \text{if dist}_{\text{Manh}} \neq 1 \\ 0 & \text{if dist}_{\text{Manh}} = 1 \end{cases} \tag{24}$$

where $\text{dist}_{\text{Manh}}$ is the normalized Manhattan distance between the compared distributions, dist_{Sp} is the difference between the specificities of the two possibility distributions, using Yager's definition of a specificity measure given in equation (5), such that: $\text{dist}_{\text{Sp}}(\pi_1, \pi_2) = |\text{Sp}(\pi_1) - \text{Sp}(\pi_2)|$, and α and β are two coefficients such that $\alpha \in [0, 1], \beta \in [0, 1]$.

In the general case, the same weight can be assigned to the two coefficients α and β. However, it is worth mentioning that the strategy to select the coefficients has a crucial impact on the results since it decides about which one of the two aspects (point-to-point, Yager's specificity) is mostly considered. Therefore, the choice of the two coefficients α and β is strongly depending on the context of application.

InfoSpecificity satisfies the properties of a measure of similarity such as described in Section III.a:

Non-negativity

i) By definition, $\mathrm{Sp}(\pi_1) \in [0,1]$, $\mathrm{Sp}(\pi_2) \in [0,1]$, then, $\mathrm{dist}_{\mathrm{Sp}}(\pi_1, \pi_2) \in [0, 1]$.

ii) $\mathrm{dist}_{\mathrm{Manh}}(\pi_1, \pi_2) \in [0, 1]$.

i) and ii) imply InfoSpecificity (π_1, π_2) assumes values on the interval [0, 1] and then it is always positive.

Symmetry

i) $\mathrm{dist}_{\mathrm{Sp}}(\pi_1, \pi_2) = \mathrm{dist}_{\mathrm{Sp}}(\pi_2, \pi_1)$.

ii) $\mathrm{dist}_{\mathrm{Manh}}(\pi_1, \pi_2) = \mathrm{dist}_{\mathrm{Manh}}(\pi_2, \pi_1)$.

i) and ii) imply

InfoSpecificity $(\pi_1, \pi_2) =$ InfoSpecificity (π_2, π_1)

Upper Bound

If $\pi_1 = \pi_2$ then,

i) $\mathrm{dist}_{\mathrm{Manh}}(\pi_1, \pi_2) = 0$

ii) $\mathrm{dist}_{\mathrm{Sp}}(\pi_1, \pi_2) = 0$

i) and ii) imply InfoSpecificity $(\pi_1, \pi_2) = 1$

Note that $\mathrm{dist}_{\mathrm{Sp}}(\pi_1, \pi_2) = 0$ is obtained, even when $\pi_1 \neq \pi_2$, when the compared distributions have the same amount of information. In this case we have $\mathrm{dist}_{\mathrm{Manh}}(\pi_1, \pi_2) \neq 0$, so the InfoSpecificity (π_1, π_2) will be different from 1. Moreover, the maximum value of InfoSpecificity measure, is reached only if the compared distributions are identically similar.

Lower Bound

i) If $\forall\, \omega_i \in \Omega, \pi_1(\omega_i) \in \{0, 1\}, \pi_2(\omega_i) \in \{0, 1\}$, then $\mathrm{dist}_{\mathrm{Manh}}(\pi_1, \pi_2) = 1$

i) implies, InfoSpecificity $(\pi_1, \pi_2) = 0$.

The InfoSpecificity achieves its lower value only when the compared distributions are perfectly contradictory.

Specificity

If $\mathrm{Sp}(\pi_1) \geq \mathrm{Sp}(\pi_2)$ and $\mathrm{Sp}(\pi_2) \geq \mathrm{Sp}(\pi_3)$, then obviously we have:

i) $\mathrm{dist}_{\mathrm{Sp}}(\pi_1, \pi_2) \leq \mathrm{dist}_{\mathrm{Sp}}(\pi_1, \pi_3)$

ii) $\mathrm{dist}_{\mathrm{Manh}}(\pi_1, \pi_2) \leq \mathrm{dist}_{\mathrm{Manh}}(\pi_1, \pi_3)$

i) and ii) imply

InfoSpecificity $(\pi_1, \pi_2) \geq$ InfoSpecificity (π_1, π_3)

IV. NUMERICAL ANALYSIS OF MORPHIC AND AMORPHIC-BASED SIMILARITY MEASURES

The shape of a distribution of possibilities informs about its informative value. Thus a shape resemblance between two possibilities distributions can lead to the deduction of the proximity between their encoded information. From this fact, we propose to group the possibilistic similarity measures into three categories: those based on the evaluation of the morphic aspect, those based on the magnitude as amorphic-based ones, and the hybrid category that combines morphic and amorphic criteria to assess similarity. The morphic category groups similarity measures based on information quantification concepts and usually rely on uncertainty measures.

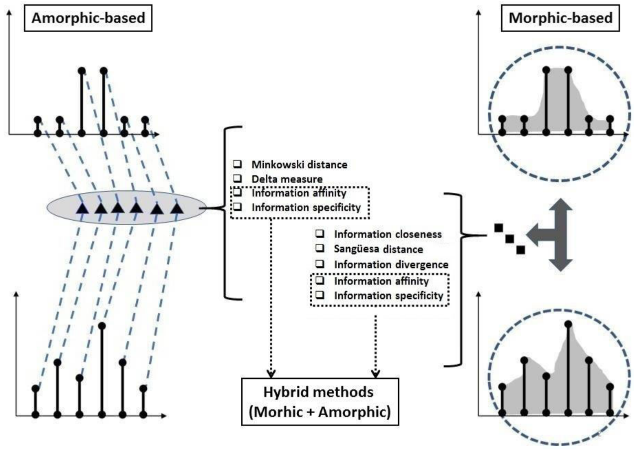


FIGURE 1. A typology of morphic and amorphic similarity measures.

For the amorphic category, it groups similarity measures according to a generalization of the metric measures. In this category, a similarity value is calculated from an estimated distance between points in a metric representation space.

Fig. 1 presents a sort of typology of the similarity measures. We note that both information affinity and information specificity measures are built considering two aspects: morphic-based and amorphic-based. Two numerical examples are hereafter considered to analyze the behavior of the similarity measures.

A. EXAMPLE 1: EVALUATION OF SIMILARITY BETWEEN TWO POSSIBILITY DISTRIBUTIONS

The first example intents to show the ability of the similarity measures to recognize the five (5) degrees of distribution resemblance labeled as: (i) Identical, (ii) Near, (iii) Informational divergent, (iv) Far and (v) Dissimilar. In fact, we consider two possibility distributions π_1 and π_2 defined over $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ as shown in Table 1. Nine measures of similarity between these two possibility distributions are evaluated versus five decreasing grades of similarity: identical, near, informational divergent, far and dissimilar. Results are given in Table 1. For InfoAffinity and InfoSpecificity measures (the last two rows), an equal weight has been assigned to both coefficients α and β.

Table 1 presents results where the measures are being calculated on specific numerical cases. It shows that the similarity measures cannot always recognize the degree of resemblance between the distributions. For example, we observe that Sangüesa distance assigns a full similarity degree to contradictory distributions (right column of Table 1). Table 2 presents in a Yes-No tabular way, if the measures of similarity respect the required properties discussed in Section III.

B. EXAMPLE 2 - EVALUATION OF SIMILARITY VERSUS THE EXTREME CASES OF CERTAINTY (COMPLETE KNOWLEDGE AND TOTAL IGNORANCE)

The possibility theory is mainly devoted to model uncertainty and to handle incomplete information. This implies that a

TABLE 1. Similarity evaluation of similarity between two distributions (π_1 *and* π_2).

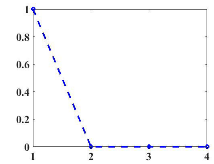
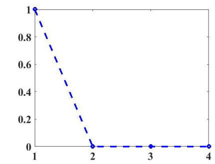
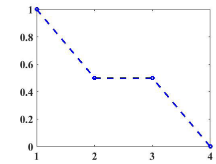
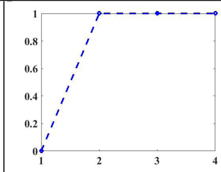
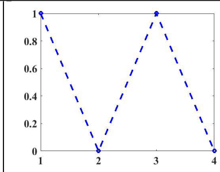
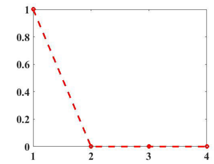
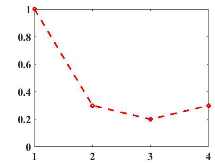
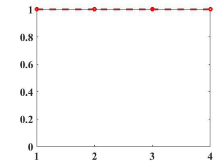
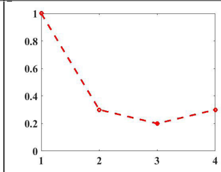
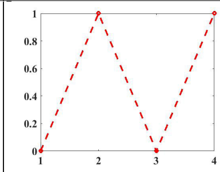
	Identical	Near	Informational divergent	Far	Contradictory
π_1					
π_2					
InfoCloseness	1	0.768	0.5	0.231	0.179
Sangüesa distance	1	0.8985	0.3962	0.25	1
Information divergence	1	1	0	0	0
Manhattan distance	1	0.8	0.5	0.2	0
Euclidean distance	1	0.78	0.5	0.19	0
Maximum distance	1	0.7	0	0	0
Delta	1	0.555	0.5	0.2	0
InfoAffinity	1	0.9	0.75	0.25	0
InfoSpecificity	1	0.7583	0.4583	0.3250	0

TABLE 2. Measures of similarity versus the six required properties: (Y=property RESPECTED; N=property not respected).

	P1: Non-negativity	P2: Symmetry	P3: Upper Bound	P4: Lower Bound	P5: Specificity	P6: Permutation
InfoCloseness	Y	Y	Y	N	Y	Y
Sangüesa distance	Y	Y	N	N	Y	Y
Information divergence	Y	N	N	N	Y	Y
Manhattan distance	Y	Y	Y	Y	Y	Y
Euclidean distance	Y	Y	Y	Y	Y	Y
Maximum distance	Y	Y	Y	Y	Y	Y
Delta	Y	Y	Y	Y	Y	Y
InfoAffinity	Y	Y	Y	Y	Y	Y
InfoSpecificity	Y	Y	Y	Y	Y	Y

similarity measure should reflect the closeness in terms of the uncertainty level. Therefore, it is expected to reach a very low similarity when comparing two distributions that are extremely different in terms of certainty.

Let π_1 and π_2 be two possibility distributions such that $\pi_1 = [0, 1]$ (i.e. complete knowledge) and $\pi_2 = [1, 1]$ (i.e. total ignorance). Table 3 shows the ability of the similarity measures to discriminate between these extreme knowledge cases.

C. RESULT ANALYSIS OF EXAMPLES 1 AND 2

This section discusses the results presented in Tables 1-3. We present our remarks in point-form notes according to the main two categories of measures: morphic and amorphic.

- Morphic based similarity measures are not always able to reveal the exact state of knowledge.
- Information Closeness doesn't respect property 4. It cannot detect contradictory distributions ($s(\pi_1, \pi_2) = 0.179 \neq 0$) as it is shown in Table 1.
- Sangüesa distance does not satisfy properties 3 and 4. It assigns a full similarity degree between two contradictory possibility distributions. In fact, it reaches its maximum value when the distribution difference represents the complete ignorance, which is the case for contradictory distributions.
- Information divergence does not satisfy property 2. It is easy to check that $D(\pi_1|\pi_2) \neq D(\pi_2|\pi_1)$. It does not respect properties 3 and 4 neither. This measure

TABLE 3. Evaluation of similarity versus the extreme cases of knowledge.

Similarity measure	$s(\pi_1, \pi_2)$
InfoCloseness	0.5
Sangüesa distance	0.6309
Information divergence	0
Manhattan distance	0.5
Euclidean distance	0 .5
Maximum distance	0.5
Delta	0.5
InfoAffinity	0.75
InfoSpecificity	0.1667

assigns a full similarity degree to distributions not similar. It also assigns the lowest similarity degree to distributions not extremely dissimilar that could adversely affect its ability to discriminate different states of knowledge.

- The amorphic-based similarity measures can deal with the basic mathematical properties of a possibilistic similarity measure (introduced in section III) since they already satisfy the properties associated to point-to-point distances (i.e. non-negativity, symmetry, identity, and triangle inequality). However, they are not able to assess two extremely different states of knowledge (i.e. complete knowledge and total ignorance) as being dissimilar ($s(\pi_1, \pi_2) = 0.5$) as it is shown in Table 3.
- Similarity based on the maximum distance does not satisfy property 4 since it assigns a null degree when comparing distributions being not contradictory (when comparing 'informational divergent' and 'far' distributions). In fact, it achieves its minimum value when the distributions share the highest degree for the same element (i. e. $\text{Inc}(\pi_1, \pi_2) = 1$).
- Hybridization of the two types (morphic and amorphic) of similarity measures is a way to enrich the expression of the degree of closeness between distributions.
- Authors in [36] have demonstrated that InfoAffinity satisfies the six basic mathematical properties of a possibilistic similarity measure. Moreover, we have demonstrated (in previous section), that InfoSpecificity also satisfies all the six properties.
- The performance of these measures in a specific context is related to the ability of their composition to assess the degree of closeness between two distributions (respectively, inconsistency measure and the specificity measure).
- InfoAffinity relies on the inconsistency measure and the metric distance. Remember that the inconsistency is the $\max(\min(\pi_1, \pi_2))$. It expresses the highest value of agreement between the two possibility distributions π_1 and π_2. The impact of the inconsistency use may make the similarity increasing more than desirable as it is shown in Tables 1 and 3 (respectively, the case of informational divergent and the extreme cases of certainty).
- InfoSpecificity uses a measure related to the variation of the amount of information between distributions, combined with a classic metric distance. Information specificity detect well whether the distributions are identical, near, informational divergent, far or contradictory. Also, referring to Table 3, it achieves the expected low value when comparing distributions that are extremely different in terms of certainty.

We recall that both InfoAffinity and InfoSpecificity rely on a coefficient that can decide whether we rely more on the morphic or the amorphic aspect. The choice of coefficients can have a significant impact on the performance of these measures. Discussion on this impact is beyond the scope of this paper.

V. EXPERIMENTAL RESULTS

This section presents results of an experimental investigation using four benchmark data sets extracted from the UCI repository [37] of machine learning databases. These databases are specifically dedicated to classification problems. The size of the databases is sufficiently large to infer a referent possibility distribution that we can consider as ground truth.

- **Base1:** Banknote authentication data sets that contains 1372 instances of 10 features extracted from images of banknotes to decide whether the banknote is genuine or forged.
- **Base2:** Page blocks classification data sets that is devoted to document analysis. It contains 5473 instances, each one describes a block of a document that can be identified according to the values of 10 attributes, in one out of 5 classes.
- **Base3:** Pima Indians diabetes data set that aims to predict diabetes disease referred on 768 individuals and based on eight factors.
- **Base4:** Statlog data set (Vehicle Silhouettes) that seeks to classify a given silhouette from 946 instances as one of four vehicles types, using a set of 18 features extracted from the silhouette.

To assess the performance of a similarity measure, we have to consider a large variety of conditions or factors that may affect the performance. These factors are:

- **Data quality:** Degraded data quality may be due to high sensor noise, the interference from other signals during acquisition or other alterations that affect the data representativeness of the modeled sample.
- **The possibility distribution estimation approach:** This influences directly the quality of the model and its ability to represent the entity.
- **The size and the number of samples:** The sample size mainly affects the quality of construction of the class model.

TABLE 4. Similarity measures considered for evaluation.

	Morphic	Amorphic	Hybrid methods (Morphic+Amorhic)
InfoCloseness	✓		
Sangüesa distance	✓		
Information divergence	✓		
Manhattan distance		✓	
Delta		✓	
InfoAffinity			✓
InfoSpecificity			✓

TABLE 5. Attribute selection results (4 databases).

Base	Base1	Base2	Base3	Base4
Number of instances	1372	5473	768	946
Number of attributes	10	10	8	18
Number of classes	2	5	2	4
Selected attribute	Kurtosis of Wavelet	Kurtosis about major axis	Percentage of black pixels within the block	Diabetes pedigree function

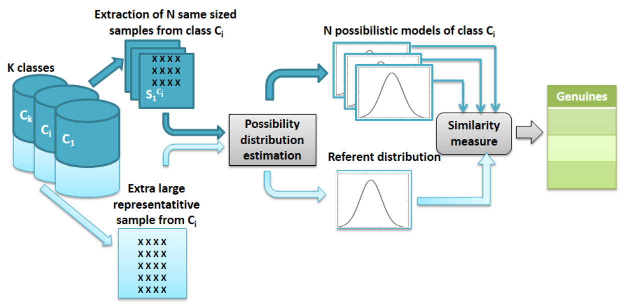


FIGURE 2. Gathering the geniuses: possibility distributions belonging to the same class.

A. DESCRIPTION OF THE EXPERIMENTAL METHODOLOGY

Our objective is to compare the measures of similarity used within a possibilistic context. Seven similarity measures are considered for experimentation. They are listed in Table 4 according to the three main categories (morphic, amorphic and hybrid).

The possibility distributions of the classes, according to one specific attribute, are being built from the four databases: Banknote authentication, Page blocks, Pima Indians diabetes, Statlog. These distributions are constructed using some well-known approaches: the symmetric transformation of Dubois-Prade [38], the asymmetric transformation of Dubois-Prade [39], and the transformation of Klir [40], [41]. Two experiments are being conducted; identified below as Scenario-1 and Scenario-2.

1) SELECTION OF AN ATTRIBUTE FOR EACH DATABASE

For each of the four databases, a unique attribute is being considered to model their respective classes. Considering the influence of the attribute on the evaluation of a similarity measure, we classify the attributes into three categories as follows:

- Non-discriminative attributes: all the similarity measures give useless results.
- Moderately discriminative attributes: an evaluation based on this type of attribute leads to a meaningful comparison between similarity measures.
- Very discriminative attributes: enable the similarity measures to easily differentiate between the samples of each class.

The selection step for the attributes has been conducted on each database using the Information Gain method [42] of the Weka application [43]. Information Gain method is an entropy-based method. It measures the relevance of a given attribute as the amount of information is gained by the acquisition of the attribute values. It was initially suggested for decision tree algorithms and afterwards, it had proven its significance for feature selection in different machine learning contexts. Other commonly used metrics, often, agree on the selection of the most pertinent attributes. In general, it is not straightforward to decide about the most efficient metric for a selected data set in a given context.

The attribute selection results are given in Table 5.

2) SCENARIO-1 DISCRIMINATION BETWEEN GENUINE (SIMILAR, POSITIVE) AND IMPOSTER (DISSIMILAR, NEGATIVE) CLASSES

In possibility theory, a similarity measure can reveal the closeness between possibility distributions. This allows, if close enough, to gather them into the same class. Moreover, it can also indicate the distance between possibility distributions to differentiate between them. When comparing two objects, a good similarity measure will give significant values while performed on objects belonging to the same class. On the other hand, it will produce small values while performed on objects belonging to a different class. That has been illustrated in Fig. 2 and Fig. 3 for separating objects belonging to different classes (imposters) and for gathering objects of the same class (genuine).

Experiments have been conducted, for each similarity measure S and for the four databases, according to the following steps.

- **Step 1 :** Different samples of the same size and belonging to the same class are extracted. Their respective possibility distributions are then matched based on the similarity measure S to a referent possibility distribution of their class. This is to build the genuine distribution (Fig. 2). Note that a good similarity measure should detect a high degree of similarity.
- **Step 2 :** Different samples of the same size are extracted from different classes. Their respective possibility distributions are then matched based on the similarity measure S to the referent distributions of all classes other

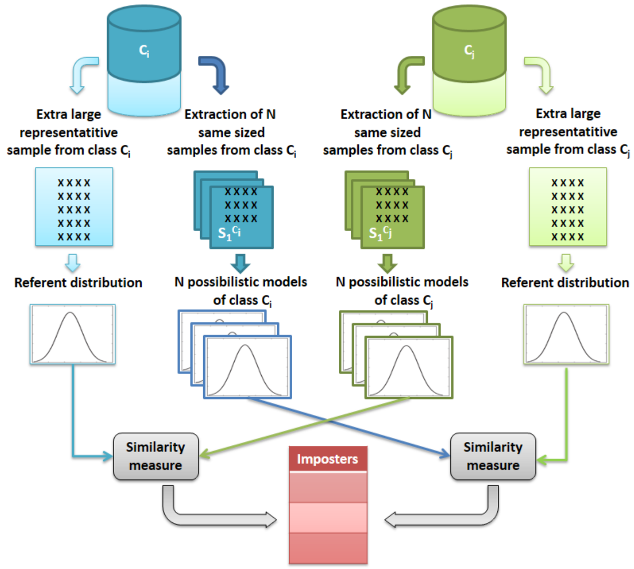


FIGURE 3. Gathering imposters: possibility distributions belonging to different classes.

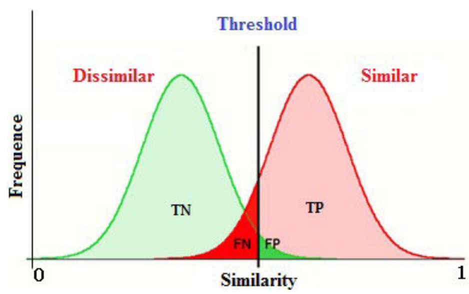


FIGURE 4. Frequency of occurrence of similarity measures in the two "Similar" and "Dissimilar" matching categories.

than their own class. This is to build the imposter distribution (Fig. 3). Note that a good similarity measure should detect a low degree of similarity.

A representation of the frequency of appearance of the similarity degrees in the two respective categories "Similar" and "Dissimilar," on a scale of increasing similarity, is given in Figure 4.

A threshold must be considered, as shown in Fig. 4, to determine if the result of a matching test has to be considered as a type "Similar" or "Dissimilar". True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) rates are then performed according to the threshold values.

In this context, we use the Receiving Operator Characteristics (ROC) curve [44] for evaluating and comparing similarity measures. This is the representation of the True Positive Rate (TP) as a function of the false positive rate (FP) as shown in Fig.5. Area under the ROC curve (AUC) quantifies the efficiency of a similarity measure to differentiate between the similarity and dissimilarity cases. AUC value is a performance criterion. A measure is considered as much efficient as it achieves a high "AUC" value.

In this evaluation, the procedure is as follows:

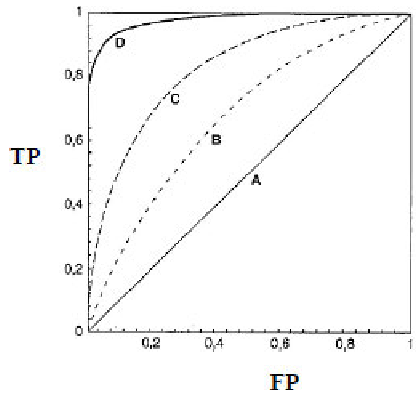


FIGURE 5. ROC curve representations: (A) No interest, (B) Bad, (C) Better than B, (D) Good approach.

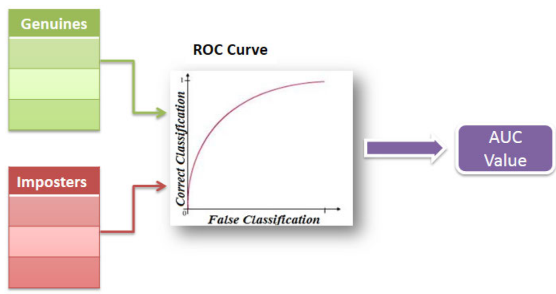


FIGURE 6. Assessment approach based on AUC value.

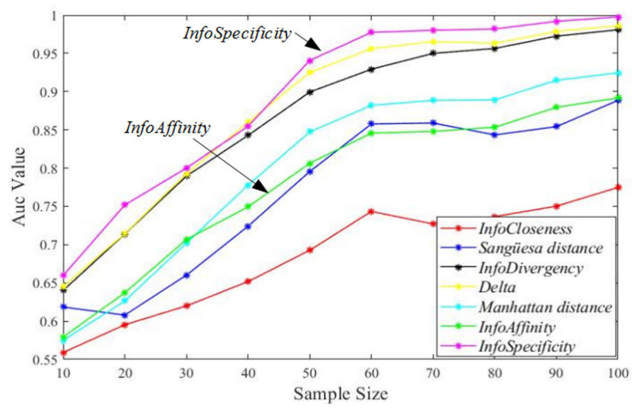


FIGURE 7. Comparison of the 7 similarity measures for Banknote Authentication data base.

- The possibility distributions are generated using the asymmetric Dubois and Prade method [39].
- The sample size mainly affects the quality of the class model construction. To get a model of good quality, we conducted the experiment for a large range of sample sizes, going from 10 to 100 elements.

Table 6 is illustrating the number of extracted samples from each data set for each considered sample size.

Fig. 7-10 illustrate, respectively for the 4 datasets, the obtained AUC values according to each considered sample size, for the seven similarity measures. The results are as follows:

TABLE 6. Number of used samples according to each data set.

Database \ Sample size	10	20	30	40	50	60	70	80	90	100
Banknote	941	920	901	881	861	841	821	801	781	761
Page Blocks	3126	3106	3086	3066	3044	3022	3002	2988	2966	2946
Diabetes	442	440	420	400	380	360	340	320	300	280
Statlog	240	220	200	180	160	140	120	100	80	60

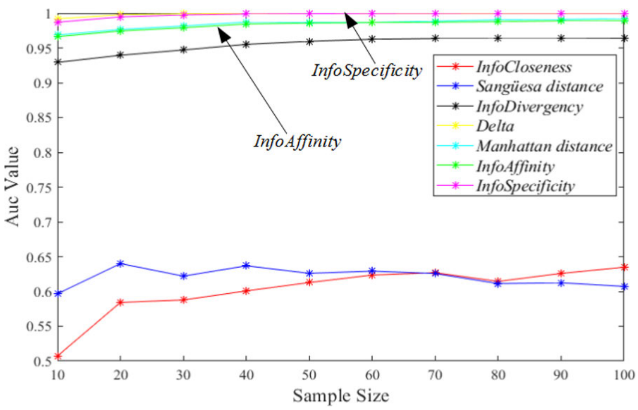


FIGURE 8. Comparison of the 7 similarity measures for Page Blocks data base.

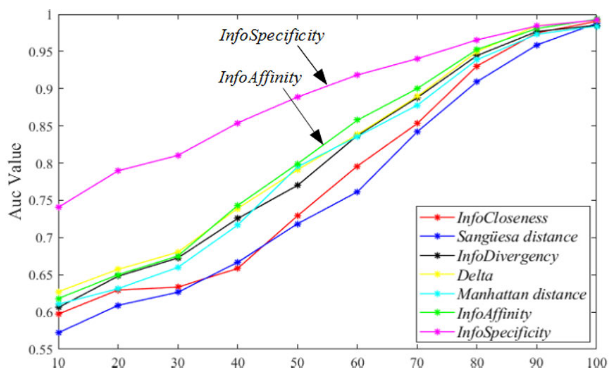


FIGURE 9. Comparison of the 7 similarity measures for Diabetes data base.

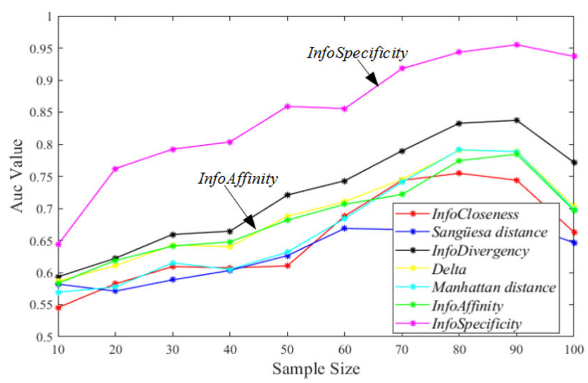


FIGURE 10. Comparison of the 7 similarity measures for Statlog data base.

- Based on the obtained AUC results, the majority of the compared measures perform well when the size of the sample increases. The results show that the measure "InfoSpecificity" (hybrid) outperforms all the other ones.

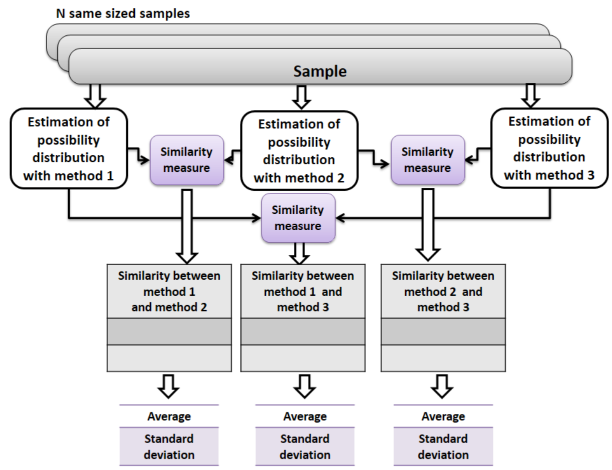


FIGURE 11. Similarity assessment of possibility distributions belonging to the same sample while estimated with different methods.

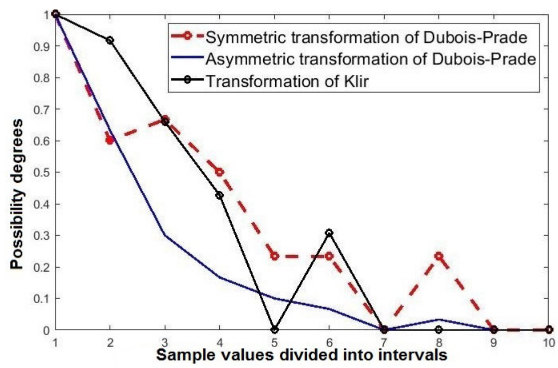


FIGURE 12. Application of different estimation methods on a sample of size=30 from Diabetes database.

- The poor results obtained with Sangüesa distance and Information divergence, especially on Page Block data set (Fig. 8), can be explained by the fact that, with the 'Sangüesa distance', two contradictory possibility distributions are considered as similar while the 'Information divergence' measure attributes a full similarity degree to distributions that are not similar and full dissimilarity degree to distributions that are not completely dissimilar.
- InfoSpecificity gives high values of similarity when compared to other similarity measures and in particular, the one based on the Manhattan distance. This is a hybrid method so it exploits both the morphic and magnitude aspects in measuring similarity.

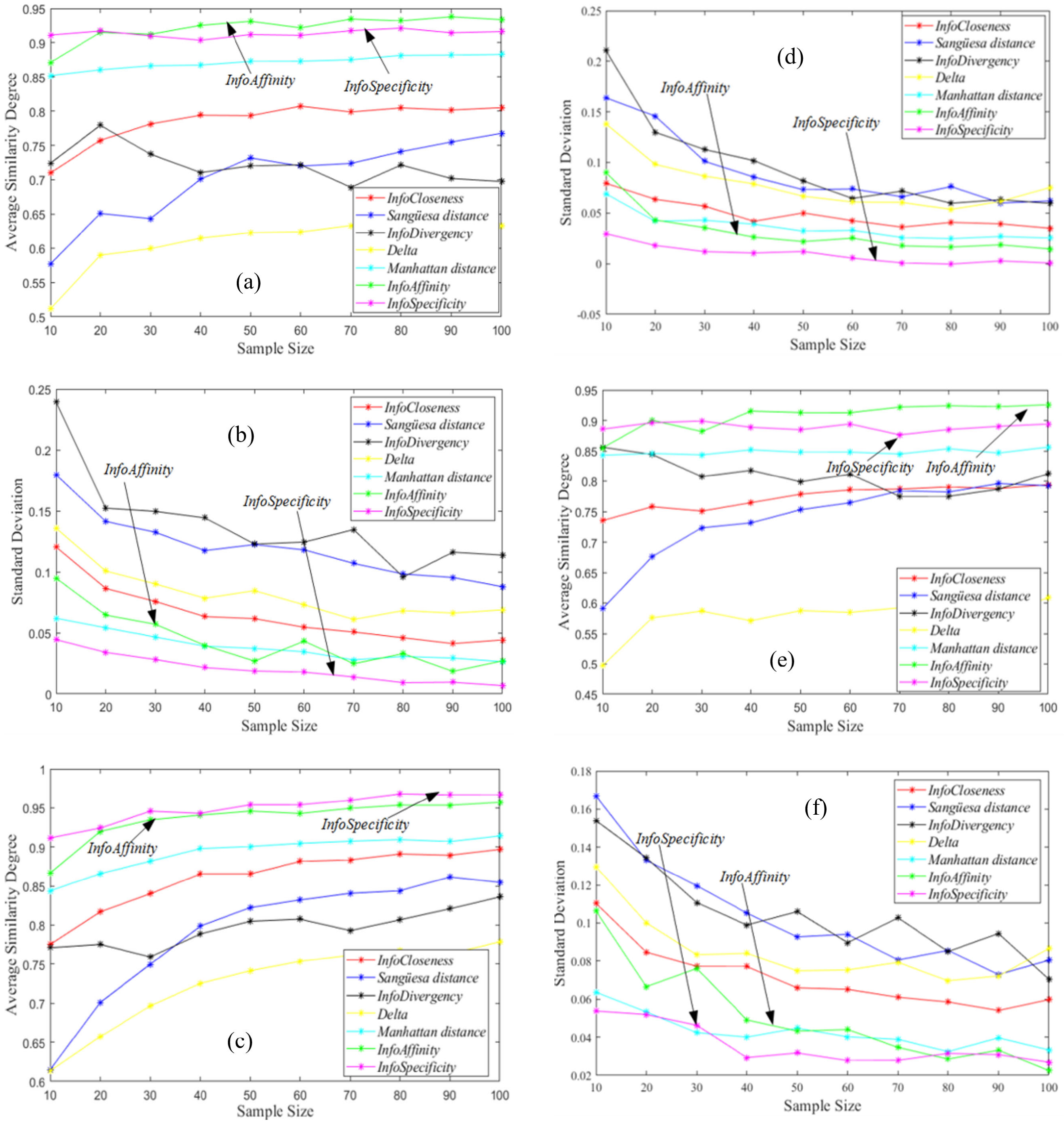


FIGURE 13. Comparison of similarity measures based on the tests of estimation methods for Diabetes database. (a, b) Comparison of Symmetric and Asymmetric transformations of Dubois-Prade based: (a) on average similarity degree and; (b) on standard deviation. (c, d) Comparison of Asymmetric transformation of Dubois-Prade and Transformation of Klir based: (c) on average similarity degree and; (d) on standard deviation. (e, f) Comparison of Symmetric transformation of Dubois-Prade and Transformation of Klir based: (e) on average similarity degree and; (f) on standard deviation.

- InfoSpecificity measure produces better results than the other measures in the context of small sample sizes. This result was expected. InfoSpecificity is based on the possibility theory which is known to be a good modeling tool in the context of information scarcity.

3) SCENARIO-2 CLOSENESS OF POSSIBILITY DISTRIBUTIONS

Scenario-2 aims at verifying if a similarity measure can reveal the closeness between distributions obtained from the same samples but estimated with different methods.

Scenario-2 consists on extracting, repeatedly, 400 different samples of the same size from Diabetes database, and applying different methods on them to estimate possibility distributions. The predicted possibility distributions of each sample are then compared. We keep the average degree and the standard deviation of similarity obtained with each measure of similarity (Fig. 11). A good similarity measure has to give a high degree of similarity indicating that the compared distributions refer to the same sample. Also, a low standard deviation indicates that a similarity measure gives almost the same results during the evaluation process.

The three estimation methods used are listed as:

1) Symmetric transformation of Dubois-Prade [38];
2) Asymmetric transformation of Dubois-Prade [39];
3) Transformation of Klir [41].

Fig. 12 illustrates three different methods to estimate the possibility distributions for the same sample. We note visually that these distributions seem to have a high degree of similarity since they have roughly a similar shape (curvature). Furthermore, they have a correspondence on the element having the maximum value. In this figure, we vary the sample size from 10 observations to 100 observations to visualize the proximity between possibility distributions with respect to the sample size.

Fig. 13 illustrates, for the Diabetes database, the similarity average degree and the standard deviation, according to each considered sample size, and for the seven similarity measures. The two criteria are applied to the set of similarity values, obtained after comparing between pairs of estimation methods.

Fig. 13 shows that the best results, either on small sample size or on big sample size, are always given by the InfoAffinity and the InfoSpecificity measures. The higher similarity values obtained with InfoAffinity and InfoSpecificity are explained mainly by their hybridization with the Manhattan distance, which has proven to be efficient in a possibilistic context.

By definition, the inconsistency of two distributions that share at least one fully possible element is equal to 1. Often, the estimated distributions from the same sample use to share the same largest element. That explains the high similarity values obtained with InfoAffinity since they are based on inconsistency. Moreover, distributions that have been inferred from the same sample, usually, share a similar distribution shape. In scenario-2, InfoSpecificity allows to detect a high similarity between the compared distributions.

VI. CONCLUSION

In this paper, we have studied the similarity between possibility distributions. First, we have analysed a set of seven similarity measures and their required properties. These measures were classified into three main categories: morphic, amorphic and hybrid. Results show that some of the existing measures give inconsistent results in conditions of data incompleteness (e.g. small sample size). To circumvent that, we proposed a new hybrid similarity measure called InfoSpecificity that uses the variation of the amount of information between distributions combined with a standard metric distance. Experiments have been conducted both on numerical examples and on four experimental databases (UCLI). The results show the importance of hybridizing morphic and amorphic approaches such as in InfoSpecificity and InfoAffinity. The hybrid methods outperformed the other methods when tested against real data. For instance, InfoSpecificity has shown its capacity to deal with small sample sizes. The performance in such special conditions proves that it is well suited for possibilistic modeling contexts that are poor-data environments.

REFERENCES

[1] K. Weick, *Sensemaking in Organizations*. Thousand Oaks, CA, USA: Sage, 1995.

[2] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *Proc. AAAI*, 2006, pp. 775–780.

[3] B. Wen-Hao, Z. An, Q. Ling-Hui, and G. Tao, "The method of measuring conflict evidence based on the modified probability distribution function and similarity measure," in *Proc. 32nd Chin. Control Conf. (CCC)*, 2013, pp. 4698–4702.

[4] L. Dengfeng and C. Chuntian, "New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions," *Pattern Recognit. Lett.*, vol. 23, nos. 1–3, pp. 221–225, Jan. 2002.

[5] L. I. Kuncheva, "Using measures of similarity and inclusion for multiple classifier fusion by decision templates," *Fuzzy Sets Syst.*, vol. 122, no. 3, pp. 401–407, Sep. 2001.

[6] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, and L. Cazzanti, "Similarity-based classification: Concepts and algorithms," *J. Mach. Learn. Res.*, vol. 10, pp. 747–776, Mar. 2009.

[7] A. A. Goshtasby, *Image Registration: Principles, Tools and Methods*. Springer, 2012.

[8] D. B. Bisandu, R. Prasad, and M. M. Liman, "Data clustering using efficient similarity measures," *J. Statist. Manage. Syst.*, vol. 22, no. 5, pp. 901–922, Jul. 2019.

[9] J. Zhao, J. Han, and L. Shao, "Unconstrained face recognition using a set-to-set distance measure on deep learned features," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2679–2689, Oct. 2018.

[10] I. Jenhani, N. B. Amor, Z. Elouedi, S. Benferhat, and K. Mellouli, "Information affinity: A new similarity measure for possibilistic uncertain information," in *Symbolic and Quantitative Approaches to Reasoning With Uncertainty*. Berlin, Germany: Springer, 2007, pp. 840–852.

[11] A. Charfi, "Evaluation de mesures de simularité en vue de leur utilisation dans des arbres de décicion dans un cadre possibiliste," M.S. thesis, Sfax Eng. School, Univ. Sfax, Sfax, Tunisia, 2014.

[12] V. Kuppili, M. Biswas, D. R. Edla, K. J. R. Prasad, and J. S. Suri, "A mechanics-based similarity measure for text classification in machine learning paradigm," *IEEE Trans. Emerg. Topics Comput. Intell.*, to be published.

[13] M. Gahegan, R. Agrawal, A. Jaiswal, J. Luo, and K. Soon, "Measures of similarity for integrating conceptual geographical knowledge: Some ideas and some questions," in *Proc. COSIT, Workshop Semantic Similarity Meas.*, 2007, pp. 1–6.

[14] R. Budiu, C. Royer, and P. Pirolli, "Modeling information scent: A comparison of LSA, PMI and GLSA similarity measures on common tests and corpora," in *Proc. RIAO*, Pittsburgh, PA, USA, 2007, pp. 1–18.

[15] M. Pelillo, *Similarity-Based Pattern Analysis and Recognition*. London, U.K.: Springer, 2013.

[16] S. Destercke, D. Dubois, and E. Chojnacki, "Unifying practical uncertainty representations—I: Generalized p-boxes," *Int. J. Approx. Reasoning*, vol. 49, no. 3, pp. 649–663, Nov. 2008.

[17] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, Jun. 1965.

[18] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[19] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 1, no. 1, pp. 3–28, Jan. 1978.

[20] C. Mencar, G. Castellano, A. M. Fanelli, and A. Bargiela, "Similarity vs. possibility in measuring fuzzy sets distinguishability," in *Proc. 5th Int. Conf. Recent Adv. Soft Comput. (RASC)*, Nottingham, U.K., 2004, pp. 354–359.
[21] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. New York, NY, USA: Plenum Press, 1988.
[22] R. R. Yager, "On the specificity of a possibility distribution," *Fuzzy Sets Syst.*, vol. 50, pp. 279–292, 1992.
[23] M. Becker and R. Laue, "A comparative survey of business process similarity measures," *Comput. Ind.*, vol. 63, no. 2, pp. 148–167, Feb. 2012.
[24] M. Higashi and G. J. Klir, "Measures of uncertainty and information based on possibility distributions," *Int. J. Gen. Syst.*, vol. 9, no. 1, pp. 43–58, Dec. 1982.
[25] M. Higashi and G. J. Klir, "On the notion of distance representing information closeness: Possibility and probability distributions," *Int. J. Gen. Syst.*, vol. 9, no. 2, pp. 103–115, Jan. 1983.
[26] R. Sangüesa, J. Cabós, and U. Cortés, "Possibilistic conditional independence: A similarity-based measure and its application to causal network learning," *Int. J. Approx. Reasoning*, vol. 18, nos. 1–2, pp. 145–167, Jan. 1998.
[27] B. Alsahwa, B. Solaiman, É. Bossé, S. Almouahed, and D. Guériot, "A method of spatial unmixing based on possibilistic similarity in soft pattern classification," *Fuzzy Inf. Eng.*, vol. 8, no. 3, pp. 295–314, Sep. 2016.
[28] Y. B. Salem, R. Idoudi, K. S. Ettabaa, K. Hamrouni, and B. Solaiman, "Ontology based possibilistic reasoning for breast cancer aided diagnosis," in *Proc. Eur., Medit., Middle Eastern Conf. Inf. Syst.*, 2017, pp. 353–366.
[29] A. Ammar and Z. Elouedi, "From crisp to soft possibilistic and rough meta-clustering of retail datasets," in *Proc. Int. Conf. Intell. Syst. Design Appl.*, 2017, pp. 485–495.
[30] A. Ammar and Z. Elouedi, "Selection of initial modes for rough possibilistic K-modes methods," in *Proc. Int. Symp. Methodol. Intell. Syst.*, 2017, pp. 77–86.
[31] I. K. Kallel, S. Almouahed, B. Solaiman, and É. Bossé, "An iterative possibilistic knowledge diffusion approach for blind medical image segmentation," *Pattern Recognit.*, vol. 78, pp. 182–197, Jun. 2018.
[32] I. K. Kallel, S. Almouahed, B. Alsahwa, and B. Solaiman, "The use of contextual spatial knowledge for low-quality image segmentation," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 9645–9665, Apr. 2019.
[33] T. Kroupa, "Application of the choquet integral to measures of information in possibility theory," *Int. J. Intell. Syst.*, vol. 21, no. 3, pp. 349–359, Mar. 2006.
[34] Y. F. Shi, L. H. He, and J. Chen, "Fuzzy pattern recognition based on symmetric fuzzy relative entropy," *Int. J. Intell. Syst. Appl.*, vol. 1, no. 1, pp. 68–75, Oct. 2009.
[35] D. Dubois and E. Hüllermeier, "Comparing probability measures using possibility theory: A notion of relative peakedness," *Int. J. Approx. Reasoning*, vol. 45, no. 2, pp. 364–385, Jul. 2007.
[36] I. Jenhani, N. B. Amor, Z. Elouedi, S. Benferhat, and K. Mellouli, "Information affinity: A new similarity measure for possibilistic uncertain information," in *Proc. Eur. Conf. Symbolic Quant. Approaches Reasoning Uncertainty*, 2007, pp. 840–852.
[37] (Mar. 5, 2019). *Pima Indians Diabetes Dataset*. [Online]. Available: http://archive.ics.uci.edu/ml/machine-learningdatabases/pima-indiansdiabetes/pima-indiansdiabetes.data
[38] D. Dubois, H. Prade, and S. Sandri, "On possibility/probability transformations," in *Fuzzy Logic*. Dordrecht, The Netherlands: Springer, 1993, pp. 103–112.
[39] D. Dubois, L. Foulloy, G. Mauris, and H. Prade, "Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities," *Reliable Comput.*, vol. 10, no. 4, pp. 273–297, Aug. 2004.
[40] G. J. Klir and J. F. Geer, "Information-preserving probability-possibility transformations," in *Fuzzy Logic*. Dordrecht, The Netherlands: Springer, 1993, pp. 417–428.
[41] G. J. Klir and B. Parviz, "Probability-possibility transformations: A comparison," *Int. J. Gen. Syst.*, vol. 21, no. 3, pp. 291–310, Aug. 1992.
[42] B. Azhagusundari and A. S. Thanamani, "Feature selection based on information gain," *Int. J. Innov. Technol. Exploring Eng.*, vol. 2, no. 2, pp. 18–21, 2013.
[43] I. H. Witten, E. Frank, L. Trigg, M. Hall, G. Holme, and S. J. Cunningham, "Weka: Practical machine learning tools and techniques with Java implementations," Univ. Waikato, Hamilton, New Zealand, Working Paper 99/11, 1999.
[44] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.

AMAL CHARFI received the Diploma degree in computer science and multimedia and the master's degree in computer science and multimedia from the Higher Institute of Computer Science and Multimedia, Sfax, in 2011 and 2014, respectively. She is currently pursuing the Ph.D. degree with the CEM Laboratory, National Engineering School of Sfax. She is also a member of the CEM Laboratory, National Engineering School of Sfax (ENIS) and of the Cybernics Team, Control and Energy Management Laboratory (CEMLab), University of Sfax. Her thesis project is included in the P.R.F project (Federated Research Programs), entitled PBMLT: Multimodal Biometric Platform for Prevention of Terrorism under the theme Beat against terrorism. Her main research interests center on uncertain information modeling, biometric recognition systems, information fusion, and machine learning.

SONDA AMMAR BOUHAMED received the Diploma degree in informatique and multimedia in 2009, the master's degree from Rouen University, France, in 2010, and the Ph.D. Degree from the National Engineering School of Sfax, Tunisia, in 2014. She is currently an Assistant Professor with the National School of Electronics and Telecommunication of Sfax (Enet-Com), a member of the Cybernics Team, Control and Energy Management Laboratory (CEMLab), University of Sfax, an Associate Researcher with the Image & Information Processing (iTi) Department, IMT Atlantique Brest, France, and a member of the Cybernics Team, University of Sfax. Her research interests range from uncertainty modeling, data and information quality, machine learning, image and signal processing, analytics and information fusion technologies to applied research in pattern recognition, smart real-time systems for obstacle detection and recognition, and decision support systems.

ÉLOI BOSSÉ (Senior Member, IEEE) received the B.A.Sc., M.Sc., and Ph.D. degrees from the Université Laval, QC, Canada, in 1979, 1981, and 1990, respectively, all in electrical engineering. In 1981, he joined the Communications Research Centre, Ottawa, Canada, where he worked on signal processing and high resolution spectral analysis. In 1988, he was transferred to the Defence Research Establishment Ottawa to work on radar target tracking in multipath. In 1992, he moved to Defence Research and Development Canada Valcartier (DRDC Valcartier) to lead a group of defense scientists on analytics, information fusion, and decision support. From 1993 to 2015, he has held an Adjunct Professor positions at several universities at the Université Laval, the University of Calgary, and McMaster University. He headed the C2 Decision Support Systems Section, DRDC Valcartier, from 1998 to 2011. He represented Canada as a DRDC member in numerous international research fora under the various cooperation research programs (NATO, TTCP, bi, and tri-laterals) in his area of expertise. He has coauthored and a Co-Editor of 5-6 books on information fusion. He retired from DRDC in September 2011. Since 2011, he conducted some research activities under NATO Peace and Security Programme, as a Researcher with the Mathematics and Industrial Engineering Department, Polytechnic of Montreal, as an Associate Researcher at IMT Atlantique, Brest, France, as a Researcher at McMaster University, Canada, and since 2015, has been the President of Expertise Parafuse, Inc., a consultant research firm on analytics and information fusion technologies. He has published more than 200 articles in journals, book chapters, conference proceedings, and technical reports. His research interests are mainly in analytics and machine learning, information fusion, decision support, and complex systems.

IMENE KHANFIR KALLEL received the Diploma degree in electrical engineering, the master's degree, and the Ph.D. degree, and the HDR degree from the National Engineering School of Sfax, Tunisia, in 2001, 2003, 2010, and 2019, respectively. She is currently an Associate Professor with the Biotechnology Higher Institute of Sfax, a member of the Control & Energy Management Laboratory (CEMLab), University of Sfax, an Associate Researcher with the Image & Information Processing (iTi) Department, IMT Atlantique, Brest, France. In 2019, she created the Cybernics Team, Sfax University, where she is working on the fusion of human, machine, and information systems. The team objective is to conduct applied research to develop human assistive technologies in biorobotics. She has been the Head of a research team working on decision support system based applications since 2011. Her research work has been mainly applied to biometrics and biotechnologies. She has published more than 30 articles in journals and conferences. Her research interests range from information theory, image and information processing, machine learning, classification, pattern recognition to analytics, and information fusion to computer-based decision support systems. She co-chaired CSP Conference of SSD Multiconference from 2015 to 2018 and actually co-chairs SCIS of SSD Multiconference. She is on the Editorial Board of the *Journal on Digital Signal Processing* (JDP) (i-manager), since February 2015.

WASSIM BOUCHAALA received the degree in electrical engineering from the National Engineers' School of Tunis, in 2003, and the master's degree in Automatic control and signal processing in 2006. After a preparatory cycle for engineering studies (mathematics & physics) and engineering studies at the National Engineers' School of Tunis, he contributed in the development of many projects in the fields of image processing, information modeling, optimization, embedded programming (PIC microcontroller), web, and mobile development. He is currently an Engineer in computer science, electronics, electrical and embedded systems, with strong skills and knowledge in mathematics and physics.

BASEL SOLAIMAN received the degree in telecommunication engineering from the Ecole Nationale Supérieure des Télécommunications de Bretagne, in 1983, and the DEA, Ph.D., and HDR degrees from the Université de Rennes-I, France, in 1983, 1988, and 1997, respectively. In 1983, he joined the Laboratoire d'Electronique de Philips (L.E.P), Paris, France, where he worked on image compression standards. In 1988, he joined the Institute of Industrial Informatics, Brest, France, as a R&D Engineer-in-Chief to work on new technologies transfer into the industrial regional tissue. In 1992, he joined the Ecole National Supérieure des Télécommunications de Bretagne (Telecom Bretagne), Brest, France, where he is currently a Full Professor and the Head of the Image and Information Processing Department. He has published several academic books, more than 180 journal articles, several book chapters, more than 230 conference proceedings, and technical reports. He has held visiting professor positions at several universities the University of Bristol, U.K.; the University of Carabobo, Venezuela; the University of Michigan, Ann Arbor, USA; the Technical University of Budapest, Hungary; E.T.S, Montreal, Canada; M.U.T, Warsaw, Poland; and the Technical University of Melaka (UTEM), Malaysia. His application domains research activities range from civilian into military ones: medical, remote sensing, underwater imaging, machine learning, information, and knowledge fusion. He chaired several international conferences focusing on his research activities and related to signal and image processing, data and information fusion, and pattern recognition.

NABIL DERBEL (Senior Member, IEEE) received the Diploma degree in engineering from the National School of Engineers of Sfax, in 1986, the master's degree (by research) in automatic control from the National Institute of Applied Sciences of Toulouse, France, in 1986, and the Ph.D. degree from the National School of Engineers of Tunis, in 1989. He joined Tunisian University, where he held different position involved in research and education. He is currently a Full Professor of automatic control with the National School of Engineers of Sfax. He has authored or coauthored more than 80 articles published in international journals and of more than 400 articles published in international conferences. His current interests include optimal control, complex systems, fuzzy logic, neural networks, and genetic algorithm.