Leila Baccour*, Adel M. Alimi and Robert I. John

Intuitionistic Fuzzy Similarity Measures and Their Role in Classification

DOI 10.1515/jisys-2015-0086
Received August 13, 2015.

Abstract: We present some similarity and distance measures between intuitionistic fuzzy sets (IFSs). Thus, we propose two semi-metric distance measures between IFSs. The measures are applied to classification of shapes and handwritten Arabic sentences described with intuitionistic fuzzy information. The experimental results permitted to do a comparative analysis between intuitionistic fuzzy similarity and distance measures, which can facilitate the selection of such measure in similar applications.

Keywords: Similarity measures, distance measures, intuitionistic fuzzy sets, semi-metric distance measures, handwritten Arabic sentence classification, shape classification, type-1 fuzzy sets.

1 Introduction

Similarity and distance measures between IFSs [1, 2] have been proposed and discussed using different approaches. Some of the proposed measures are generalizations of crisp or fuzzy distance measures. Atanassov [2] and Szmidt and Kacprzyk [39] generalized crisp Hamming and Euclidean distance measures to IFSs, and Hung and Yang [24, 25] generalized Hausdorff distance and L_p metric to propose similarity measures between IFSs. Hung and Yang [26] generalized similarity measures between type-1 fuzzy sets to propose intuitionistic fuzzy similarity measures (IFSMs).

The existing intuitionistic fuzzy similarity measures (IFSMs) in literature are examined and compared [8, 29], and many measures are proposed to resolve deficiencies of existing measures. For instance, Hong and Kim [22] proposed IFSMs to resolve deficiencies found in the measures presented in Refs. [14, 15], but these IFSMs have some counter-intuitive cases resolved by Fan and Zhangian [20] who proposed another IFSM. Szmidt and Kacprzyk [40] proposed IFSMs using the complementary of intuitionistic fuzzy sets to measure the dissimilarity between IFSs in addition to their similarity. More details about IFSMs are presented in Ref. [11].

In general, the existing IFSMs are demonstrated to validate properties. However, some of them do not validate one or more properties, and others are shown to have some counter-intuitive cases. A comparative analysis by Li et al. [29] shows the problems found in some IFSMs from literature, and the relationship between some IFSMs are done in Ref. [10].

In crisp logic, a distance measure, which does not validate triangular inequality, is considered as semi-metric distance measure (see Appendices). Some measures between intuitionistic fuzzy sets, in literature, do not validate the fourth property of similarity or distance measure. Could these measures be used in operations of classification? Is the satisfaction of the fourth property mandatory to obtain good results in classification? Through analogy to crisp distance measures, we propose semi-metric distance

***Corresponding author: Leila Baccour,** REGIM-Lab.: Research Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia, e-mail: leilabaccour@gmail.com
Adel M. Alimi: REGIM-Lab.: Research Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax 3038, Tunisia
Robert I. John: Automated Scheduling, Optimisation and Planning (ASAP) group, Room C72 Computer Science Jubilee Campus Wollaton Road Nottingham, NG8 1BB, UK

measures, and we apply them to classification to answer the precedent questions. Intuitionistic fuzzy similarity measures are crucial operators for comparison between two patterns and are used in many systems such as recognition systems, classification systems, etc.; as examples, applications include handwritten Arabic sentences recognition [5, 9], logic programming [3, 4], decision-making problems [38], and medical diagnosis [27, 37]. Despite the numerous works on IFSMs and the different propositions of similarity and distance measures, most of them did numerical applications on some artificial samples and did not use a large data set to show the validation of IFSMs for classifying patterns or another domain in which similarity can influence results. Hung and Yang [26] applied some IFSMs to students' evaluation, but the used data set was not large. A comparative analysis between IFSMS in practical and in real applications is needed. This is the objective of the paper in which we apply IFSMs to two applications of classifications using large data sets.

In Section 2, we present the definition of IFSs followed by distance and similarity measures from the literature. In Section 4, we propose two semi-metric distance measures, and we apply them to some numerical examples. Therefore, we apply them and the measures from literature to the classification of shapes and recognition of handwritten Arabic sentences. The set of sentences is extracted from the IFN/ENIT data set (set A) [36], and the shapes represent the SQUID data set. The extraction process of intuitionistic fuzzy features of the two data sets is described in Sections 5 and 6, respectively. The results of the classification obtained with all measures are compared in Sections 5 and 6.

2 Intuitionistic Fuzzy Sets

An intuitionistic fuzzy set A (IFS) in a universe of discourse X is defined by Ref. [1] as:

$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\}$, where $\mu_A(x): X \to [0,1]$ and $\nu_A(x): X \to [0,1]$ with the condition $0 \le \mu_A(x) + \nu_A(x) \le 1 \; \forall x \in X$. IFSs(X) is the set of all IFSs in X.

The numbers $\mu_A(x)$, $\nu_A(x)$ denote, respectively, the degrees of membership and of non-membership of the element $x \in X$ to A. Also, μ_A and ν_A are called, respectively, truth membership function and false membership function.

In addition to membership and non-membership functions, a function of hesitancy or uncertainty of x to A denoted by $\pi_A(x)$ must be taken into consideration. $\pi_A(x)$ is computed as:

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \text{ with } 0 \le \pi_A(x) \le 1$$

3 Similarity, Distance, and Semi-metric Measures Between IFSs

An intuitionistic fuzzy similarity measure $S(A, B)$ serves to match between two IFSs A and B of $IFSs(X)$ and should satisfy the following properties

$(P1)\ 0 \le S(A, B) \le 1$

$(P2)\ S(A, B) = 1$ if and only if $A = B$

$(P3)\ S(A, B) = S(B, A)$

$(P4)\ S(A, C) \le S(A, B)$ *and* $S(A, C) \le S(B, C)$ *if* $A \subseteq B \subseteq C$. $C \in IFSs(X)$

Note that similarity and distance (dissimilarity) measures are complementary; when the first increases, the second decreases. Normalized distance measure and similarity measure are dual concepts. So, $S(A, B) = 1 - d(A, B)$ and vice versa. The properties of distance measures below are complementary to those of similarity measures:

$(DP1)\ 0 \le d(A, B) \le 1$
$(DP2)\ d(A, B) = 0$ if and only if $A = B$
$(DP3)\ d(A, B) = d(B, A)$
$(DP4)$ If $A \subseteq B \subseteq C,\ A,\ B,\ C \in IFSs(X)$, then $d(A, C) \ge d(A, B)$ and $d(A, C) \ge d(B, C)$

In crisp logic, distance measures are classified to distance, index of distance, index of dissimilarity, semi-metric distance, or distance ultra-metric (see Appendices) according to their validations of the whole or some of properties. For IFSs, with reference and analogy to crisp distance, measures that do not validate the property *DP4* of distance measures can also be considered if they show correct classification, and thus, they will be called semi-metric distance measures.

Notation The following notations are used to simplify the formulas of distance and similarity measures:

$\mu_A^i = \mu_A(x_i)$ and $\mu_B^i = \mu_B(x_i)$
$\nu_A^i = \nu_A(x_i)$ and $\nu_B^i = \nu_B(x_i)$
$\Delta_\mu^i = \mu_A^i - \mu_B^i,\ \Delta_\nu^i = \nu_A^i - \nu_B^i$ and $\Delta_\pi^i = \pi_A^i - \pi_B^i$,

In the sequel, we present existing similarity and distance measures in Tables 1 and 2 between tow IFSs *A* and *B* of *n* elements.

3.1 Similarity Measures Between Intuitionistic Fuzzy Sets

3.1.1 Comments on Measures From Literature

- The measures $S_{C_1}^p$ (23) and $S_{C_2}^p$ (24) can be correct only if each of the sum is divided by *n* (second part of the formulas after 1-).
- In the measures $S_{C_1}^p$ and $S_{C_2}^p$, it is very difficult to determine the weight for *n* elements especially if the sum of the weights is one. So, the weight could be eliminated in these measures, that it will be possible to compute them.
- $S_{C_1}^p = d_{H_{SK}}$ when $\delta_1 = \delta_2 = \delta_3$ and when ω_i is omitted.
- S_p cannot classify patterns when they are described with both fuzzy and intuitionistic fuzzy information.
- The measure S_{pk3} (13) does not satisfy the property (P4) of similarity measures between IFSs.

Proof 1 *Let $A \subseteq B \subseteq C$, then $\mu_A(x) \le \mu_B(x) \le \mu_C(x)$ and $\nu_A(x) \ge \nu_B(x) \ge \nu_C(x)$. Therefore,*

$|\mu_A(x) - \mu_B(x)| \le |\mu_A(x) - \mu_C(x)|,\ |\nu_A(x) - \nu_B(x)| \le |\nu_A(x) - \nu_C(x)|,$
$|\mu_A(x) + \mu_B(x)| \le |\mu_A(x) + \mu_C(x)|$ and $|\nu_A(x) + \nu_B(x)| \ge |\nu_A(x) + \nu_C(x)|$

According to the above hypothesis, we have:

$|\mu_A(x) - \mu_B(x)| + |\nu_A(x) - \nu_B(x)| \le |\mu_A(x) - \mu_C(x)| + |\nu_A(x) - \nu_C(x)|$ but the inequality
$|\mu_A(x) + \mu_B(x)| + |\nu_A(x) + \nu_B(x)| \ge |\mu_A(x) + \mu_C(x)| + |\nu_A(x) + \nu_C(x)|$ is not true.

Consequently, this inequality is not true, too:

$$1 - \frac{|\mu_A(x) - \mu_B(x)|}{|\mu_A(x) + \mu_B(x)|} + \frac{|\nu_A(x) - \nu_B(x)|}{|\nu_A(x) + \nu_B(x)|} \ge 1 - \frac{|\mu_A(x) - \mu_C(x)|}{|\mu_A(x) + \mu_C(x)|} + \frac{|\nu_A(x) - \nu_C(x)|}{|\nu_A(x) + \nu_C(x)|},$$

and then $S(A, B) \ge S(A, C)$ is not true. Thus, the property P4 is not verified for the similarity measure S_{pk3}.

Other comments about similarity measures and their relationships are given in Ref. [10].

Table 1: Similarity Measures Between IFSs.

Authors	IFSMs	
Chen [14, 15]	$S_C(A,B)=1-\frac{\sum_{i=1}^{n}\lvert S_A(x_i)-S_B(x_i)\rvert}{2n}$ where $S_A(x_i)=\mu_A(x_i)-\nu_A(x_i)$ and $S_B(x_i)=\mu_B(x_i)-\nu_B(x_i)$.	(1)
Hong and Kim [22]	$S_H(A,B)=1-\frac{\sum_{i=1}^{n}\lvert\Delta_\mu^i\rvert+\lvert\Delta_\nu^i\rvert}{2n}$.	(2)
Fan and Zhangian [20]	$S_L(A,B)=1-\frac{\sum_{i=1}^{n}\lvert S_A(x_i)-S_B(x_i)\rvert}{4n}-\frac{\sum_{i=1}^{n}\lvert\Delta_\mu^i\rvert+\lvert\Delta_\nu^i\rvert}{4n}$	(3)
Dengfeng and Chuntian [19]	$S_d^p(A,B)=1-\frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}\lvert\varphi_A(x_i)-\varphi_B(x_i)\rvert^p}$ $\varphi_A(x_i)=\frac{\mu_A(x_i)+1-\nu_A(x_i)}{2}$ and $\varphi_B(x_i)=\frac{\mu_B(x_i)+1-\nu_B(x_i)}{2}$ with $1\le p<+\infty$	(4)
Mitchell [32]	$S_{mod}(A,B)=\frac{1}{2}(\varphi_\mu(A,B)+\varphi_\Phi(A,B))$ where $\varphi_\mu(A,B)$ and $\varphi_\Phi(A,B)$ denote, respectively, the similarity measures between the low membership functions μ_A and μ_B and the high membership functions $\Phi_A=1-\nu_A$ and $\Phi_B=1-\nu_B$ as follows: $\varphi_\mu(A,B)=S_d^p(\mu_A(x_i),\mu_B(x_i))=1-\frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}\lvert\Delta_\mu^i\rvert^p}$ $\varphi_\Phi(A,B)=S_d^p(\Phi_A(x_i),\Phi_B(x_i))=1-\frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}\lvert\Delta_\nu^i\rvert^p}$	(5)
Liang and Shi [28]	$S_e^p(A,B)=1-\frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}(\varphi_{\mu_{AB}}(x_i)+\varphi_{\nu_{AB}}(x_i))^p}$ where $\varphi_{\mu_{AB}}(x_i)=\frac{\lvert\Delta_\mu^i\rvert}{2}$ and $\varphi_{\nu_{AB}}(x_i)=\left\lvert\frac{1-\nu_A(x_i)}{2}-\frac{1-\nu_B(x_i)}{2}\right\rvert$	(6)
	$S_s^p(A,B)=1-\frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}(\varphi_{s_1}(x_i)+\varphi_{s_2}(x_i))^p}$ with $\varphi_{s_1}(x_i)=\frac{\lvert m_{A_1}(x_i)-m_{B_1}(x_i)\rvert}{2}$ and $\varphi_{s_2}(x_i)=\frac{\lvert m_{A_2}(x_i)-m_{B_2}(x_i)\rvert}{2}$ where A_1 and A_2 are the sub-intervals from A divided by the median value denoted $m_A(x_i)=\frac{\mu_A(x_i)+1-\nu_A(x_i)}{2}$ of the interval $[\mu_A(x_i), 1-\nu_A(x_i)]$. The median values of A_1 and A_2 are $m_{A_1}(x_i)=\frac{m_A(x_i)+\mu_A(x_i)}{2}$ and $m_{A_2}(x_i)=\frac{m_A(x_i)+1-\nu_A(x_i)}{2}$. The median values $m_{B_1}(x_i)$ and $m_{B_2}(x_i)$ of B_1 and B_2 are computed similarly.	(7)
	$S_h^p(A,B)=1-\frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}\left(\sum_{m=1}^{3}\omega_m\varphi_m(x_i)\right)^p}$ where: $0\le\omega_m\le1$, $\sum_{m=1}^{3}\omega_m=1$, $\varphi_1(x_i)=\varphi_{s_1}(x_i)+\varphi_{s_2}(x_i)$, $\varphi_2(x_i)=\lvert\varphi_A(x_i)-\varphi_B(x_i)\rvert$, $\varphi_A(x_i)$ and $\varphi_B(x_i)$ are defined in formula (4). $\varphi_3(x_i)=\max\left(\frac{\pi_A(x_i)}{2},\frac{\pi_B(x_i)}{2}\right)-\min\left(\frac{\pi_A(x_i)}{2},\frac{\pi_B(x_i)}{2}\right)$.	(8)
Zhang and Fu [44]	$S_{ZF}(A,B)=1-\frac{1}{2n}\sum_{i=1}^{n}(\lvert\delta_A(x_i)-\delta_B(x_i)\rvert+\lvert\alpha_A(x_i)-\alpha_B(x_i)\rvert)$ where: $\delta_A(x_i)=\mu_A(x_i)+\pi_A(x_i)\mu_A(x)$ and $\alpha_A(x_i)=\nu_A(x_i)+\pi_A(x_i)\nu_A(x_i)$	(9)
Hung and Yang [26]	$S_{w1}(A,B)=\frac{1}{n}\sum_{i=1}^{n}\frac{\min(\mu_A(x_i),\mu_B(x_i))+min(\nu_A(x_i),\nu_B(x_i))}{\max(\mu_A(x_i),\mu_B(x_i))+\max(\nu_A(x_i),\nu_B(x_i))}$	(10)

Table 1 (continued)

Authors	IFSMs
	$S_{pk1}(A, B)=\frac{\sum_{i=1}^{n}\min(\mu_A(x_i), \mu_B(x_i))+\min(\nu_A(x_i), \nu_B(x_i))}{\sum_{i=1}^{n}\max(\mu_A(x_i), \mu_B(x_i))+max(\nu_A(x_i), \nu_B(x_i))}$ (11)
	$S_{pk2}(A, B)=1-\frac{1}{2}(\max_i\lvert\mu_A(x_i)-\mu_B(x_i)\rvert+\max_i\lvert\nu_A(x_i)-\nu_B(x_i)\rvert)$ (12)
	$S_{pk3}(A, B)=1-\frac{\sum_{i=1}^{n}\lvert\mu_A(x_i)-\mu_B(x_i)\rvert+\lvert\nu_A(x_i)-\nu_B(x_i)\rvert}{\sum_{i=1}^{n}(\lvert\mu_A(x_i)+\mu_B(x_i)\rvert+\lvert\nu_A(x_i)+\nu_B(x_i)\rvert)}$ (13)
	$S_{new1}(A, B)=1-\frac{1-\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\lvert\mu_A(x_i)-\mu_B(x_i)\rvert+\lvert\nu_A(x_i)-\nu_B(x_i)\rvert\right)}{1-\exp(-n)}$ (14)
	$S_{new2}(A, B)=1-\frac{1-\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\lvert\sqrt{\mu_A(x_i)}-\sqrt{\mu_B(x_i)}\rvert+\lvert\sqrt{\nu_A(x_i)}-\sqrt{\nu_B(x_i)}\rvert\right)}{1-\exp(-n)}$ (15)
Szmidt and Kacprzyk [40]	Similarity measures between elements *x* and *y*: $Sim_1(x, y)=1-f(d_{H_{SK}}(x, y), d_{H_{SK}}(x, y^c))$ (16) $Sim_2(x, y)=\frac{1-f(d_{H_{SK}}(x, y), d_{H_{SK}}(x, y^c))}{1+f(d_{H_{SK}}(x, y), d_{H_{SK}}(x, y^c))}$ (17) $Sim_3(x, y)=\frac{1-f(d_{H_{SK}}(x, y), d_{H_{SK}}(x, y^c))^2}{1+f(d_{H_{SK}}(x, y),d_{H_{SK}}(x, y^c))^2}$ (18) $Sim_4(x, y)=\frac{\exp(-f(d_{H_{SK}}(x, y), d_{H_{SK}}(x, y^c)))-\exp(-1)}{1-\exp(-1)}$ (19) where: $f(d_{H_{SK}}(x, y), d_{H_{SK}}(x, y^c))=\frac{d_{H_{SK}}(x, y)}{d_{H_{SK}}(x, y)+d_{H_{SK}}(x, y^c)}$ and y^c the complement of *y* and $d_{H_{SK}}$ (28) is defined in Section 3.2. Thus, a measure between two IFSs *A* and *B* is defined as: $Sim_k(A, B)=\frac{1}{n}\sum_{i=1}^{n}sim_k(d_{H_{SK}}(x_i, y_i), d_{H_{SK}}(x_i, y_i^c))$, for $k=1\ldots4$ (20)
Ye [43]	Cosine similarity measure between IFSs: $S_Y(A, B)=\frac{1}{n}\sum_{i=1}^{n}\frac{\mu_A(x_i)\mu_B(x_i)+\nu_A(x_i)\nu_B(x_i)}{\sqrt{\mu_A^2(x_i)+\nu_A^2(x_i)}\sqrt{\mu_B^2(x_i)+\nu_B^2(x_i)}}$ (21)
Park et al. [35]	$S_P=1-\frac{1}{n}\sum_{i=1}^{n}\lvert\Delta_\pi^i\rvert$ (22)
Chu et al. [16]	$S_{C_1}^p=1-\sum_{i=1}^{n}w_i(\delta_1\lvert\Delta_\mu^i\rvert^p+\delta_2\lvert\Delta_\nu^i\rvert^p+\delta_3\lvert\Delta_\pi^i\rvert^p)^{\frac{1}{p}}$ (23) where $\sum_{j=1}^{3}\delta_j=1,\ 0<\delta_j<1$ and $\sum_{i=1}^{n}w_i=1$ $S_{C_2}^p=1-\left[\sum_{i=1}^{n}w_j\left(\delta_1\left\{\frac{\lvert\Delta_\mu^i\rvert}{\max(\mu_A^i, \mu_B^i)}\right\}^p+\delta_2\left\{\frac{\lvert\Delta_\nu^i\rvert}{\max(\nu_A^i, \nu_B^i)}\right\}^p+\delta_3\left\{\frac{\lvert\Delta_\pi^i\rvert}{\max(\pi_A^i, \pi_B^i)}\right\}^p\right)\right]^{\frac{1}{p}}$ (24) It is considered that $\frac{0}{0}=0$
Mukherjee and Basu [33]	$S_{MB}=\frac{\sum_{i=1}^{n}(\mu_A^i\mu_B^i+\nu_A^i\nu_B^i+\pi_A^i\pi_B^i)}{\max\left(\sum_{i=1}^{n}\left(\mu_A^{i^2}+\nu_A^{i^2}+\pi_A^{i^2}\right), \sum_{i=1}^{n}\left(\mu_B^{i^2}+\nu_B^{i^2}+\pi_B^{i^2}\right)\right)}$ (25)

Table 2: Distance Measures Between IFSs.

Authors	Distance measures	
Atanassov [2]	Hamming and Euclidean distances	
	$d_{H_A}(A, B) = \frac{1}{2n}\sum_{i=1}^{n} \lvert\Delta_\mu^i\rvert + \lvert\Delta_\nu^i\rvert$	(26)
	$d_{E_A}(A, B) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} (\Delta_\mu^i)^2 + (\Delta_\nu^i)^2}.$	(27)
Szmidt and Kacprzyk [39]	Hamming and Euclidean distances	
	$d_{H_{SK}}(A, B) = \frac{1}{2n}\sum_{i=1}^{n} \lvert\Delta_\mu^i\rvert + \lvert\Delta_\nu^i\rvert + \lvert\Delta_\pi^i\rvert.$	(28)
	$d_{E_{SK}}(A, B) = \frac{1}{\sqrt{2n}}\sqrt{\sum_{i=1}^{n} (\Delta_\mu^i)^2 + (\Delta_\nu^i)^2 + (\Delta_\pi^i)^2}.$	(29)
Hung and Yang [24]	Hausdorff distance	
	$d_{Hd}(A, B) = \frac{1}{n}\sum_{i=1}^{n} \max(\lvert\Delta_\mu^i\rvert, \lvert\Delta_\nu^i\rvert)$	(30)
	Similarity measures using $d_{Hd}(A, B)$:	
	$S_{HY_1}(A, B) = 1 - d_{Hd}(A, B)$	(31)
	$S_{HY_2}(A, B) = \frac{\exp^{-d_{Hd}(A,B)} - \exp^{-1}}{1 - \exp^{-1}}$	(32)
	$S_{HY_3}(A, B) = \frac{1 - d_{Hd}(A, B)}{1 + d_{Hd}(A, B)}$	(33)
Wang and Xin [42]	$d_{WX_1}(A, B) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\lvert\Delta_\mu^i\rvert + \lvert\Delta_\nu^i\rvert}{4}\right] + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\max(\lvert\Delta_\mu^i\rvert, \lvert\Delta_\nu^i\rvert)}{2}\right]$	(34)
	$d_{WX_2}^p(A, B) = \frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n} (\varphi_\mu(x_i) + \varphi_\nu(x_i))^p}$	(35)
	where $\varphi_\mu(x_i) = \lvert\mu_A(x_i) - \mu_B(x_i)\rvert/2$, $\varphi_\nu(x_i) = \lvert\nu_A(x_i) - \nu_B(x_i)\rvert/2$ and $p > 0$	
Huang et al. [23]	$d_{H_1}^p(A, B) = \left(\sum_{i=1}^{n} \frac{\lvert\Delta_\mu^i\rvert^p + \lvert\Delta_\nu^i\rvert^p}{2n}\right)^{\frac{1}{P}}$	(36)
	$d_{H_2}(A, B) = \frac{1}{n}\sum_{i=1}^{n} \lambda_1 \max(\lvert\Delta_\mu^i\rvert, \lvert\Delta_\nu^i\rvert) + \lambda_2 \min(\lvert\Delta_\mu^i\rvert, \lvert\Delta_\nu^i\rvert)$	(37)
	$d_{H_3}^p(A, B) = \sqrt[p]{\frac{1}{n}\sum_{i=1}^{n} \lambda_1 \max(\lvert\Delta_\mu^i\rvert^p, \lvert\Delta_\nu^i\rvert^p) + \lambda_2 \min(\lvert\Delta_\mu^i\rvert^p, \lvert\Delta_\nu^i\rvert^p)}$	(38)
	where $0 \le \lambda_1 \le 1$, $0 \le \lambda_2 \le 1$, $\lambda_1 \ge \lambda_2$, $\lambda_1 + \lambda_2 = 1$	
	If $\lambda_1 = 1, \lambda_2 = 0$ or $\lambda_1 = \lambda_2 = \frac{1}{2}$, $p = 1$ the d_{H_2} and d_{H_3} distances lead to the Hamming distances d_{HA} and d_{HSK}, respectively. Having the same hypothesis and for $p = 2$, the distance d_{H_3} leads to Euclidean distance d_{E_A}	
	$d_{H_4}(A, B) = \frac{1}{n}\sum_{i=1}^{n} \frac{2(\lvert\Delta_\mu^i\rvert + \lvert\Delta_\nu^i\rvert)}{2 + \lvert\Delta_\mu^i\rvert + \lvert\Delta_\nu^i\rvert}$	(39)
	$d_{H_5}(A, B) = \frac{2\sum_{i=1}^{n}(\lvert\Delta_\mu^i\rvert + \lvert\Delta_\nu^i\rvert)}{2n + \sum_{i=1}^{n}\lvert\Delta_\mu^i\rvert + \lvert\Delta_\nu^i\rvert}$	(40)
	$d_{H_6}(A, B) = \frac{2d_{H_1}(A, B)}{1 + d_{H_4}(A, B)}$	(41)
Hung and Yang [25]	The following IFSMs are proposed based on L_p metric	
	$S_{HY_4}^p(A, B) = \frac{2^{\frac{1}{p}} - L_p(A, B)}{2^{\frac{1}{p}}}$	(42)

Table 2 (continued)

Authors	Distance measures	
	$S^p_{HY_5}(A, B) = \frac{\exp(-L_p(A, B)) - \exp(-2^{\frac{1}{p}})}{1 - \exp(-2^{\frac{1}{p}})}$	(43)
	$S^p_{HY_6}(A, B) = \frac{2^{\frac{1}{p}} - L_p(A, B)}{2^{\frac{1}{p}}(1 + L_p(A, B))}$	(44)
	where: $L_p(A, B) = \frac{1}{n}\sum_{i=1}^{n} d_p(I_A, I_B)$	
	$d_p(I_A(x_i), I_B(x_i)) = (\lvert\Delta^i_\mu\rvert^p + \lvert\Delta^i_\nu\rvert^p)^{1/p}$, $p \geq 1$	
	and $I_A(x_i) = [\mu_A(x_i), 1 - \nu_A(x_i)]$ and $I_B(x_i) = [\mu_B(x_i), 1 - \nu_B(x_i)]$, $i = 1, 2, \ldots, n$	
Boran and Akay [13]	$d^p_{BA}(A, B) = \sqrt[p]{\frac{1}{2n(t+1)^p}\sum_{i=1}^{n}\lvert t*\Delta^i_\mu - \Delta^i_\nu\rvert^p + \lvert t*\Delta^i_\nu - \Delta^i_\mu\rvert^p}$	(45)
	where $p = 1, 2, 3, \ldots$ and $t = 2, 3, 4, \ldots$ identifies the level of uncertainty	

3.2 Distance Measures Between Intuitionistic Fuzzy Sets

Some distance measures are proposed based on crisp distances, and some similarity measure propositions are based on distance measures. We cite the following measures in Table 2 from literature.

4 Proposition of Semi-Metric Distance Measures

In the following, we propose two semi-metric distance measures between IFSs. The first semi-metric distance measure is defined as:

$$d^p_{SM}(A, B) = \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{\lvert\mu_A(x_i) - \mu_B(x_i)\rvert^p}{\mu_A(x_i) + \mu_B(x_i)} + \frac{\lvert\nu_A(x_i) - \nu_B(x_i)\rvert^p}{2 - \nu_A(x_i) - \nu_B(x_i)}\right). \tag{46}$$

where $p \geq 1$. If $\mu_A(x_i) + \mu_B(x_i) = 0$ or $2 - \nu_A(x_i) - \nu_B(x_i) = 0$, then, for the i^{th} value x_i, $d^p_{SM}(A, B) = 0$.

It is easy to remark that d^p_{SM} satisfies the properties *DP*1–*DP*3. Next, we verify if it satisfies the property *DP*4.

Proof 2 *Let $A \subseteq B \subseteq C$, then, $\mu_A(x_i) \leq \mu_B(x_i) \leq \mu_C(x_i)$ and $\nu_A(x_i) \geq \nu_B(x_i) \geq \nu_C(x_i)$. Therefore,*

$\lvert\mu_A(x_i) - \mu_B(x_i)\rvert \leq \lvert\mu_A(x_i) - \mu_C(x_i)\rvert$, $\lvert\nu_A(x_i) - \nu_B(x_i)\rvert \leq \lvert\nu_A(x_i) - \nu_C(x_i)\rvert$,
$\lvert\mu_A(x_i) + \mu_B(x_i)\rvert \leq \lvert\mu_A(x_i) + \mu_C(x_i)\rvert$, and $\lvert 2 - \nu_A(x_i) - \nu_B(x_i)\rvert \geq \lvert 2 - \nu_A(x_i) - \nu_C(x_i)\rvert$

as $p \geq 0$, we have:

$\lvert\mu_A(x_i) - \mu_B(x_i)\rvert^p \leq \lvert\mu_A(x_i) - \mu_C(x_i)\rvert^p$, $\lvert\nu_A(x_i) - \nu_B(x_i)\rvert^p \leq \lvert\nu_A(x_i) - \nu_C(x_i)\rvert^p$.

According to the above hypothesis, we have:

$$\frac{\lvert\nu_A(x_i) - \nu_B(x_i)\rvert^p}{\lvert 2 - \nu_A(x_i) + \nu_B(x_i)\rvert} \leq \frac{\lvert\nu_A(x_i) - \nu_C(x_i)\rvert^p}{\lvert 2 - \nu_A(x_i) + \nu_C(x_i)\rvert},$$

and the inequality: $\frac{|\mu_A(x_i)-\mu_B(x_i)|^p}{|\mu_A(x_i)+\mu_B(x_i)|} \le \frac{|\mu_A(x_i)-\mu_C(x_i)|^p}{|\mu_A(x_i)+\mu_C(x_i)|}$ *is not true. Consequently, this inequality is not true, too:*

$$\frac{|\mu_A(x_i)-\mu_B(x_i)|^p}{|\mu_A(x_i)+\mu_B(x_i)|}+\frac{|\nu_A(x_i)-\nu_B(x_i)|^p}{|2-\nu_A(x_i)-\nu_B(x_i)|} \le \frac{|\mu_A(x_i)-\mu_C(x_i)|^p}{|\mu_A(x_i)+\mu_C(x_i)|}+\frac{|\nu_A(x_i)-\nu_C(x_i)|^p}{|2-\nu_A(x_i)-\nu_C(x_i)|}.$$

So $d_{SM}^p(A, B) \le d_{SM}^p(A, C)$ *is not true. Thus, the property (DP4) is not verified for the distance measure* d_{SM}^p.

The second semi-metric distance measure is extending a similarity measure defined between two type-1 fuzzy sets A' and B' in Ref. [34] as:

$$L(A', B') = 1 - \max_i(|\mu_{A'}(x_i) - \mu_{B'}(x_i)|). \tag{47}$$

Based on the measure $L(A', B')$, we propose the following measure:

$$d_L(A, B) = (\max_i(|\mu_A(x_i)-\mu_B(x_i)|, |\nu_A(x_i)-\nu_B(x_i)|) + \max_i(|\pi_A(x_i)-\pi_B(x_i)|))/2. \tag{48}$$

The distance d_L is a distance that satisfies the properties *DP*1–*DP*3, but it does not satisfy the property *DP*4 of distance measures between IFSs as shown by the following proof:

Proof 3 *$d_L(A, C) \ge d_L(A, B)$ but $d_L(A, C) \ge d_L(B, C)$ is not always verified. $d_L(A, C) = \max_i(|\mu_A(x_i)-\mu_C(x_i)|, |\nu_A(x_i)-\nu_C(x_i)|)$, $d_L(B, C) = \max_i(|\mu_B(x_i)-\mu_C(x_i)|, |\nu_B(x_i)-\nu_C(x_i)|)$. The maximum can be $|\mu_A(x_i)-\mu_B(x_i)|$ and $|\mu_B(x_i)-\mu_C(x_i)|$ or $|\nu_A(x_i)-\nu_B(x_i)|$ and $|\nu_B(x_i)-\nu_C(x_i)|$ or $|\mu_A(x_i)-\mu_B(x_i)|$ and $|\nu_B(x_i)-\nu_C(x_i)|$ or $|\nu_A(x_i)-\nu_B(x_i)|$ and $|\mu_B(x_i)-\mu_C(x_i)|$. In the two first cases, $d_L(A, C) \ge d_L(B, C)$, but in the two last cases, $d_L(A, C) \ge d_L(B, C)$ is not always true. Taking into account the second term $|\pi_A - \pi_B|$, the inequality $d_L(A, C) \ge d_L(B, C)$ can be not true, too. So, the property DP4 is not verified by the distance d_L. So, it is a semi-metric distance measure.*

4.1 Numerical Examples

In this section, we apply the proposed similarity and semi-metric distance measures to some numerical examples in which patterns are borrowed from Ref. [28]. The results of comparisons between the patterns A_i, $i = 1, 2, 3$ and B of the two data sets are given in Table 3.

Example 1

Let $A_1 = \{(x_1, 0.3, 0.3), (x_2, 0.2, 0.2), (x_3, 0.1, 0.1)\}$

$A_2 = \{(x_1, 0.2, 0.2), (x_2, 0.2, 0.2), (x_3, 0.2, 0.2)\}$

$A_3 = \{(x_1, 0.4, 0.4), (x_2, 0.4, 0.4), (x_3, 0.4, 0.4)\}$

$B = \{(x_1, 0.3, 0.3), (x_2, 0.2, 0.2), (x_3, 0.1, 0.1)\}$

From this data, it is clear that $B = A_1$, so the sample B belongs to the pattern A_1.

Table 3: Numerical Examples and Comparisons Between Proposed Distance Measures.

IFSM		Example 1			Example 2	
	A_1, B	A_2, B	A_3, B	A_1, B	A_2, B	A_3, B
$d_{SM}^{p=1}$	0.0	0.110	0.250	0.381	0.018	0.539
$d_{SM}^{p=2}$	0.0	0.011	0.06	0.065	0.006	0.238
d_L	0.0	0.15	0.35	0.45	0.15	0.199

Example 2

Let $A_1 = \{(x_1, 0.1, 0.1), (x_2, 0.5, 0.1), (x_3, 0.1, 0.9)\}$

$A_2 = \{(x_1, 0.5, 0.5), (x_2, 0.7, 0.3), (x_3, 0.0, 0.8)\}$

$A_3 = \{(x_1, 0.7, 0.2), (x_2, 0.1, 0.8), (x_3, 0.4, 0.4)\}$

$B = \{(x_1, 0.4, 0.4), (x_2, 0.6, 0.2), (x_3, 0.0, 0.8)\}$

According to results of Ref. [24], sample B belongs to the pattern A_2. On examining the table below, we remark that all the proposed measures classify patterns correctly.

5 Application of Similarity Measures to Shape Classification

5.1 Data Set Presentation

We used the SQUID data set (http://www.ee.surrey.ac.uk/CVSSP/demos/css/demo.html), which contains about 1100 images of edges of fishes. Every image is represented with an extended curvature scale space (CSS) descriptors [8, 30, 31] of original shape to identify concavities of the shapes' edges and CSS descriptors of dual shapes to identify convexities of shapes; for more details, see Ref. [12]. In addition to the extended CSS descriptors, two features are used: the eccentricity and the frequency of codes in Freeman Chain Code [21] of shapes' edges. In conclusion, the features used to describe the shapes are:

- one value of eccentricity
- a vector of concavity points (ordinates points)
- a vector of convexity points (ordinates points)
- Freeman chain code description

There are 30 classes of shapes, and every shape is labeled with the cluster to which it belongs. The data set of the shapes is divided into two data sets: The first is constituted of 738 shapes that serves as reference data set, and the second is constituted of 362 shapes that serves as test data set.

5.2 Constitution of Intuitionistic Fuzzy Features of Shapes

In this section, we are interested in creating intuitionistic fuzzy features of shapes. This stage allows us to pass from a crisp data set constituted with crisp values to an intuitionistic fuzzy data set. The construction of this data set is described in the next paragraphs.

5.2.1 Constitution of Intuitionistic Fuzzy Values of Freeman Chain Code

We extract Freeman chain codes for every shape edge. Freeman codes are a succession of numbers from zero to seven. The numbers indicate directions of a pixel constructing the connected compound edge. This directions can be vertical, horizontal, or representing a curvature. The codes 1, 3, 5, and 7 indicate the pixel belongingness to a curvature.

The number (N) of each direction in the chain code is computed. Then, the frequency F of each direction is computed as N divided by the total number (NT_1) of codes indicating curvatures. Otherwise, N is divided by the total number (NT_2) of codes not indicating a curvature. So, we obtain eight information. It is obvious that all obtained information are in the interval [0,1] (N is lower than NT_1 or NT_2).

We consider two intuitionistic fuzzy sets A and B in the discourse universe $X = \{0, 1, 2, 3, 4, 5, 6, 7\}$ with $A = \{1, 3, 5, 7\}$ and $B = \{0, 2, 4, 6\}$. Set A concerns codes indicating curvature in Freeman chain, and set B concerns codes not indicating curvatures. Thus, membership and non-membership degrees are computed for two sets A and B according to the following functions:

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\} \quad B = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\}$$

$$\text{where} \begin{cases} \mu_A(x) = N / NT_1 \\ \nu_A(x) = 0 \end{cases} \qquad \textit{where} \begin{cases} \mu_B(x) = 0 \\ \nu_B(x) = N / NT_2 \end{cases}$$

5.2.2 Constitution of Intuitionistic Fuzzy Values of Eccentricity and CSS Descriptors

The CSS descriptors and the eccentricity are first fuzzified according to type-1 fuzzy sets (see [12]). The obtained results serve to constitute membership and non-membership degrees using these formulas proposed in Ref. [41].

Let $A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\}$ an IFS where μ_A and ν_A are defined by these formulas:

$$\begin{cases} \mu_A(x) = (1 - \mu_{A'}(x))^{\lambda} \\ \nu_A(x) = \lambda * \mu_{A'}(x). \end{cases} \tag{49}$$

where $\lambda \in [0,1]$ and A' a type-1 fuzzy set. In our application $\lambda = 0.8$.

The process of representation with intuitionistic fuzzy information of the features' eccentricity and CSS descriptors is described in Figure 1.

5.2.3 Experimentation and Results

We applied the proposed semi-metric distance measures and similarity measures from literature to classify the data test of shapes using KNN classifier. The process of classification is represented in Figure 2, and the

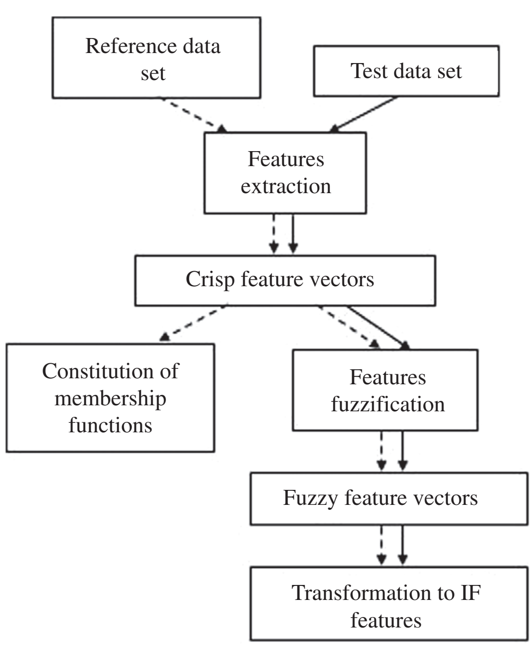

Figure 1: Process of Representation with IF Information of the Feature Eccentricity and CSS Descriptors.

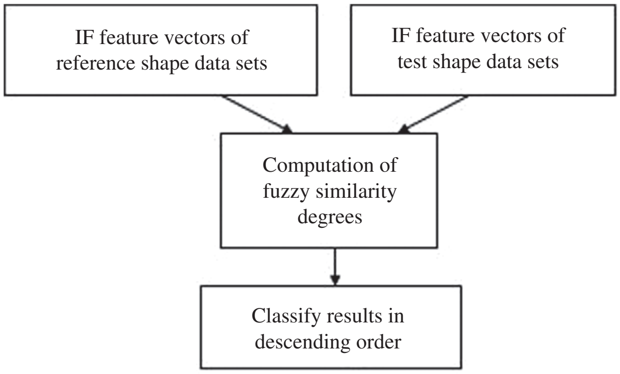

Figure 2: Classification of Shapes Using an Intuitionistic Fuzzy Similarity Measure.

results are given in Table 4 below. The 1-best presents the shape number found in the first position, and the 10-best is the sum of found shapes in the i^{th} position where $i \in 1 \ldots 10$.

In Table 4, results are classified in descending order of 10-best. The results are very close to one another, varying from 83.43% to 77.62% with a difference of 5.81%. The best results are given by S_Y followed by those of $d_{SM}^{p=2}$, d_{E_A}, $d_{E_{SK}}$, S_{MB}, $S_e^{p=1}$, $d_{WX_2}^{p=1}$ with slight differences. The lowest results are obtained by sim_k given after those of $d_{SM}^{p=1}$, S_{w1}, $S_{HY_6}^{p=1}$. The remaining measures give results between 81.77% and 79.28% with a difference of 1.74%. The results on 1-best are also very close situated between 24.86% produced by $d_{SM}^{p=1}$ and 19.34% given by sim_k with a difference of 5.52%. Similarly, results in 2-best are close with sim_k in the last rank and S_Y in the first rank. The classification results give an idea of measure applications to a large data set. Unfortunately, the results of the classification are not very high for all measures, so the used features may not fit the data set of shapes. In addition, the results given by the similarity measures are close, which puts all IFSMs in the same level with a bit of difference. We note that $S_{C_2}^{p=1}$ (24) is computed in the shape classification and Arabic sentence recognition with omission of ω. In the second section, we apply IFSMs to Arabic sentence classification, and we compare the results to have more ideas about the results of IFSMs.

6 Application of Similarity Measures to Arabic Sentence Recognition

Handwritten Arabic words present many difficulties for recognition because of the variations in information between different writers and the overlapping between the word characters. So, features representing the words cannot be exact; they only contain imprecision. This last grows when the words are not normalized. In our work, we use an extract of the IFN/ENIT data set [36] constituted of different names of Tunisian towns. The names of the towns can be formed with one word or more; so, they are considered as sentences. In this section, we describe a data set of Arabic sentences with intuitionistic fuzzy information. Thus, we apply fuzzy similarity measures from literature and the proposed semi-metric distance measures to recognize Arabic sentences [5, 9]. In the following subsections, we describe these processes with more details.

6.1 Data Set Sentence Presentation

We use 6537 images of 824 handwritten Tunisian town/village names extracted from the IFN/ENIT data set [36]. The data set images are written by different writers and undergo some preprocessing like noise reduction

Table 4: Shape Classification Results Obtained with Intuitionistic Fuzzy Similarity Measures.

IFSM	1-best (%)	2-best (%)	10-best (%)	Error rate
S_Y (21)	22.93	41.71	83.43	16.57
$d_{SM}^{p=2}$ (46)	21.82	40.1	82.32	17.68
d_{E_A} (27)	21.55	40.88	82.32	17.68
$d_{E_{SK}}$ (29)	21.54	40.33	82.32	17.68
S_{MB} (25)	21	36.19	82.04	17.96
$S_e^{p=1}$ (6), $d_{WX_2}^{p=1}$ (35)	22.65	39.78	82.22	18.78
d_L (48)	20.99	33.98	81.77	18.23
S_P (22)	23.2	37.85	81.49	18.51
S_{mod} (5), $S_s^{p=1}$ (7), $S_d^{p=1}$ (4)	22.65	39.78	81.22	18.78
$S_h^{p=1}$ (8)	22.65	39.5	81.22	18.78
$d_{BA}^{p=1}$ (45)	22.65	39.78	81.22	18.78
S_{pk2} (12)	18.23	34.25	80.94	19.06
S_{new1} (14)	22.93	38.67	80.66	19.34
S_{pk3} (13)	23.48	40.61	80.39	19.61
S_{pk1} (11)	23.2	40.88	80.39	19.61
S_{ZF} (9)	23.2	40.88	80.39	19.61
d_{WX_1} (34), $d_{H_{SH}}$ (28)	22.93	40.61	80.39	19.61
S_{HY_1} (31)	22.93	40.61	80.39	19.61
S_{HY_3} (33)	22.38	40.33	80.11	19.89
$S_{HY_4}^{p=1}$ (42), d_{H_A} (26), S_H (2)	23.48	40.61	80.11	19.89
$S_{HY_5}^{p=1}$ (43)	23.2	40.88	79.83	20.17
S_{HY_2} (32)	23.2	40.61	79.83	20.17
S_{new2} (15)	22.1	37.57	79.83	20.17
d_{H_5} (40)	18.78	38.4	79.56	20.44
S_L (3)	22.93	40.88	79.28	20.72
$d_{SM}^{p=1}$ (46)	24.86	39.23	78.73	21.27
S_{w1} (10)	22.65	40.06	78.73	21.27
$S_{HY_6}^{p=1}$ (44)	22.38	41.16	78.73	21.27
d_{H_4} (39)	22.65	39.78	78.45	21.55
$S_{C_2}^{p=1}$ (24)	20.17	35.08	77.9	22.1
Sim_k (20)	19.34	38.4	77.62	22.38

but are not normalized. The data set is divided into two data sets: reference data set constituted of 4357 sentence images and test data set constituted of 2180 images of sentences [5, 6, 12]. Recognition process is done into two steps:

- description of images by features using intuitionistic fuzzy sets
- comparison between test data set and reference data set using an intuitionistic fuzzy similarity measure

The features are first constituted according to crisp sets from word connected compounds and are:

- sum of pixel distances from the base or the top lines
- higher black pixel coordinates
- direction frequency of Freeman chain code

Thus, the features are fuzzified according to type-1 fuzzy sets, then, according to intuitionistic fuzzy sets. The descriptions of crisp and fuzzy features are presented in detail in Ref. [12]. In the following, we describe the process of constitution of intuitionistic fuzzy features.

6.1.1 Sum of Pixel Distances From the Base or the Top Lines

For the black pixel in the edge of connected compounds, three sums of mathematical distances are computed, d_1, d_2, and d_3, as described in Ref. [12]. The sum of the distances of the black pixels is divided by the sum of all black pixels of connected compound edge denoted as S. So, we obtain the distance sums as $d1=\frac{d1}{s}$, $d2=\frac{d2}{s}$ and $d3=\frac{d3}{s}$. These results are fuzzified by computing their membership degrees to trapezoidal functions. To obtain intuitionistic information, the formulas of Ref. [41] given in Section 5.2.1 is used, to compute membership and non-membership degrees as:

$$\begin{cases} \mu_A(x)=(1-\mu_{A'}(x))^{\lambda} \\ \nu_A(x)=\lambda*\mu_{A'}(x). \end{cases} \quad (50)$$

where $\lambda\in[0,1]$ (in this application, $\lambda=0.8$; this value is chosen to minimize the value of π) and A' a type-1 fuzzy set.
The membership and non-membership degrees are computed for each sum of distances.

6.1.2 Higher Black Pixel Coordinates

The abscissa of the higher black pixel can indicate its position (at the beginning, in the middle, or at the end) in the connected compound, and the ordinate comparing with its position on the top line can indicate a degree of membership to a triangular function, representing type-1 fuzzy set, to be a stroke. Three type-1 fuzzy sets are then obtained. Consequently, three intuitionistic fuzzy sets are defined and have the following description:

$$A=\{\langle x,\mu_A(x),\nu_A(x)\rangle | x\in X\}$$
$$\text{with}\,\mu_A(x)=\mu_{A'}(x)\ \text{and}\ \nu_A(x)=1-\mu_A(x)$$

where $\mu_{A'}$ is the membership degree of x to a type-1 fuzzy set represented by a triangular membership.

For each connected compound, six degrees of belongingness and non-belongingness are retained. The two first degrees represent the membership degrees of a pixel in the beginning of a connected compound, the two second degrees represent the membership degrees of a pixel in the middle of a connected compound, and the two last degrees represent the membership degrees of a pixel at the end of a connected compound.

6.1.3 Direction Frequency of Freeman Chain Code

Connected compounds are described by chain codes of Freeman. The process of computation of membership and non-membership degrees is the same presented in the Section 5.2.1.

6.2 Experimentation and Results

The similarity measures from the literature and the proposed semi-metric measure are applied to match between test data set and reference data set of Arabic sentences using the KNN classifier. The recognition results for each applied IFSM are presented in Table 5. The 1-best represents the rate of sentences found in the first position, and the 10-best is the rate of the sum of found sentences in the i^{th} position where $i\in 1\ldots 10$.

The obtained results are classified in descending order of the 10-best (see Table 5). $d_{SM}^{p=1}$ is in the first rank with recognition results equal to 91.88%, and S_{new2} is in the last rank with results of 46.23%. These

Table 5: Arabic Sentence Recognition Results Obtained with Intuitionistic Fuzzy Similarity Measures.

IFSM	1-best (%)	2-best (%)	10-best (%)	Error rate (%)
$d_{SM}^{p=1}$ (46)	67.98	77.11	91.88	8.11
$d_{SM}^{p=2}$ (46)	69.5	78.03	91.47	8.53
$S_{C_2}^{p=1}$ (24)	68.17	77.06	91.38	8.62
S_{mod} (5)	67.38	76.88	91.19	8.81
$S_e^{p=1}$ (6), $d_{WX_2}^{p=1}$ (35)	66.61	76.56	91	9
$d_{BA}^{p=1}$ (45)	67.16	76.65	90.92	9.08
$S_d^{p=1}$ (4)	66.61	76.19	91	9
S_{pk3} (13)	67.16	76.69	90.83	9.17
$S_s^{p=1}$ (7)	67.2	76.79	90.78	9.22
S_{w1} (10)	68.66	78.11	90.64	9.35
$S_h^{p=1}$ (8)	66.51	75.77	90.46	9.54
S_{ZF} (9)	67.8	76.51	90.46	9.54
$S_{HY_5}^{p=1}$ (43)	66.56	75.73	90.32	9.68
$S_{HY_6}^{p=1}$ (44)	67.11	76.19	90.32	9.68
$S_{HY_4}^{p=1}$ (42), d_{HA} (26), S_H (2)	66.28	75.41	90.32	9.68
S_L (3)	65.04	75.41	89.68	10.32
d_{WX_1} (34)	66.06	74.59	89.45	10.55
d_{E_A} (27)	63.94	73.53	89.22	10.78
S_{HY_3} (33)	66.19	74.45	89.17	10.83
S_{HY_2} (32)	66.06	74.4	88.94	11.06
S_{pk1} (11)	67.52	75.91	88.67	11.33
S_{HY_1} (31)	65.64	74.04	88.53	11.47
S_{pk2} (12)	63.21	72.8	88.17	11.83
$d_{H_{SK}}$ (28)	64.86	73.58	88.07	11.93
$d_{E_{SK}}$ (29)	64.22	72.39	87.89	12.11
d_L (48)	58.58	68.76	84.91	15.09
S_P (22)	57.94	67.11	84.59	15.41
Sim_k (20)	48.39	60.27	84.58	15.51
S_Y (21)	45.87	55.83	79.95	20.05
S_{new1} (14)	52.48	57.48	62.29	37.7
S_{MB} (25)	22.06	31.83	61.19	38.81
d_{H_4} (39)	22.48	29.4	59.22	40.78
d_{H_5} (40)	20.87	27.8	58.17	41.83
S_{new2} (15)	40.05	43.76	46.23	53.76

results show an important difference of 45.65% between the best and worst results, which can explain the differences between the results of the IFSMs. $d_{SM}^{p=1}$, S_{mod}, S_e^1, $d_{WX_2}^{p=1}$, and S_d^1 are in the second rank with bit differences in their results followed by those of S_{pk3}, S_s^p, S_{w1}, S_h^p, S, $S_{HY_5}^{p=1}$, $S_{HY_5}^{p=1}$, $S_{HY_4}^{p=1}$, d_{HA}, and S_H with results varying from 90.83% to 90.32%. The results decrease with measures d_L, sim_k and continue decreasing with measures S_Y and S_{new1}. The lowest results are given by d_{H4}, d_{H5}, S_{new2}. We note that S_e^p is proposed to resolve weakness of S_d^p, but their results are the same when $p = 1$. The IFSMs d_{HA} and d_{EA} produce better results than d_{HSK} and d_{ESK}, respectively, despite that the latter are proposed to rectify d_{HA} and d_{EA}.

In addition, some measures represent some counter-intuitive cases such as S_{mod} and S_d^p and do not verify some properties; however, they produce the best results. Similarly, the semi-metric d_{SM}^p do not satisfy the fourth property of distance measure; however, it produces the best results when $p = 1$ and $p = 2$. In the same way, the measure S_{pk3} does not verify the fourth property, but it produces acceptable results. The proposed measures d_L produces modest results. In 1-best and 2-best, the results differ slightly from the 10-best; d_{SM}^{p} and S_{w1} are in the first rank, and d_{H4}, d_{H5}, S_{new2} are in the last rank.

7 Conclusion

We presented intuitionistic fuzzy similarity and distance measures from the literature. We proposed two semi-metric distance measures. Thus, we applied them and measures from the literature to shape the classification and Arabic sentence recognition. In both applications, most measures produced acceptable results, the proposed measures d_{SM}^{p} gave the best results when $p=1$ and $p=2$ in handwritten Arabic sentence classification. d_{H4}, d_{H5}, sim_k produced low results in both applications, S_Y produced the best results in the shape classification and low results in the handwritten Arabic sentence classification. S_{new1} and S_{new2} produced low results in the handwritten Arabic sentence classification and acceptable results in the shape classification. $d_{SM}^{p=1}$ produced modest results in the first application with best results in the 1-best and best results in the second application. d_L produced modest results in both applications.

From the obtained results, we can conclude that the validation of the fourth property of similarity or distance measures has no influence on the results. Some of the measures produced best results in both applications such as $d_{SM}^{p=2}$; consequently, it must produce good results in similar applications. Similarly, d_{H4}, d_{H5}, and sim_k produced the worst results in both applications, so, they can produce low results in similar applications. The remaining similarities produced acceptable results in both applications.

The experimental results showed that a semi-metric distance measure can produce good results. In addition, some IFSMs are proposed to rectify or correct existing IFSMs, produced inferior results to those of these measures. To produce good results, an IFSM may not depend on some properties and may contain some counter-intuitive cases. IFSMs are numerous; some classification algorithms need IFSM, and an application of any IFSM may affect results. Thus, an IFSM, which satisfies all properties, does not imply that it produces the best results. The choice should only be based on experimentation. In future work, the proposed measures will be implemented on other data sets and other research topics [7, 17, 18] to assess their results.

Acknowledgments: The authors would like to acknowledge the financial support of this work by grants from the General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

Appendices

Properties of Crisp Distance Measures

A distance measure between two crisp vectors x and y, in the discourse universe E, is defined as a function $d: E^2 \to \mathbb{R}^+$ and have the following properties:

1. $\forall x \in E,\ d(x, x) = 0$ (minimality)
2. $\forall x, y \in E,\ d(x, y) = 0 \Rightarrow x = y$ (identity)
3. $\forall x, y \in E,\ d(x, y) = d(y, x)$ (symmetry)
4. $\forall x, y, z \in E,\ d(x, y) \le d(x, z) + d(z, y)$ (triangular inequality)

According to the properties, a function d is called distance, index of distance, index of dissimilarity, semi-metric distance, or distance ultra-metric as:

– sign of dissimilarity: if the function d satisfies the properties 1 and 2
– index of distance or semi-metric distance: if the function d satisfies the properties 1, 2, and 3

- distance: if the function d satisfies the properties 1, 2, 3, and 4
- distance ultra-metric: if the function d verifies the property:
  $\forall x, y, z \in E,\ d(x, y) \leq \max(d(x, z), d(z, y))$

Bibliography

[1] K. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets Syst.* **20** (1986), 87–96.
[2] K. Atanassov, *Intuitionistic fuzzy sets: theory and applications*, Physica-Verlag, Heidelberg, 1999.
[3] K. Atanassov and G. Gargov, Intuitionistic fuzzy logic, *Comptes Rendus de l'Academie des Bulgare Sciences* **43** (1990), 9–12.
[4] K. Atanassov and C. Georgeiv, Intuitionistic fuzzy prolog, *Fuzzy Sets Syst.* **53** (1993), 121–128.
[5] L. Baccour and A. M. Alimi, A comparison of some intuitionistic fuzzy similarity measures applied to handwritten Arabic sentences recognition, in: *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'2009*, pp. 1389–1392, ICC Jeju Island, Korea, 2009.
[6] L. Baccour and A. M. Alimi, Applications and comparisons of fuzzy similarity measures, in: *Proc. IEEE World Congress on Computational Intelligence, WCCI 2010, FUZZ-IEEE 2010*, pp. 1–7, Barcelone, 2010.
[7] L. Baccour and R. I. John, Experimental analysis of crisp similarity and distance measures, in: *6th International Conference on Soft Computing and Pattern Recognition, SoCPaR 2014*, pp. 96–100, Tunis, Tunisia, August 8–11, 2014.
[8] L. Baccour, W. Maghrebi, S. Kanoun and A. M. Alimi, A comparative study of a fuzzy similarity measure aggregated with fuzzy implications applied to shape recognition, in: *Proc. 6th Mexican International Conference on Artificial Intelligence MICAI 2007*, Aguascalientes-Mexico, 2007.
[9] L. Baccour, S. Kanoun, V. Maergner and A. M. Alimi, An application of intuitionistic fuzzy information for handwritten Arabic word recognition, in *Proc. Notes on IFS, the Twelfth International Conference on Intuitionistic Fuzzy Sets ICIFS'2008*, vol. 14, pp. 67–72, Sofia Bulgaria, 2008.
[10] L. Baccour, A. M. Alimi and R. I. John, Relationship between intuitionistic fuzzy similarity measures, in: *Proc. IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'2011*, pp. 971–975, Taipei, Taiwan, 2011.
[11] L. Baccour, A. M. Alimi and R. I. John, Similarity measures for intuitionistic fuzzy sets: state of the art, *J. Intell. Fuzzy Syst.* **24** (2013), 37–49.
[12] L. Baccour, A. M. Alimi and R. I. John, Some notes on fuzzy similarity measures and application to classification of shapes, recognition of Arabic sentences and mosaic, *IAENG Int. J. Comput. Sci.* **41** (2014), 81–90.
[13] F. E. Boran and D. Akay, A biparametric similarity measure on intuitionistic fuzzy sets with applications to pattern recognition, *Inform. Sci.* **255** (2014), 45–57.
[14] S. M. Chen, Measures of similarity between vague sets, *Fuzzy Sets Syst.* **74** (1995), 217–223.
[15] S. M. Chen, Similarity measures between vague sets and between elements, *IEEE Trans. Syst. Man Cybernet*, **27** (1997), 153–158.
[16] C. H. Chu, K.C. Hung and P. Julian, A complete pattern recognition approach under Atanassov's intuitionistic fuzzy sets, *Knowl. Based Syst.* **66** (2014), 36–45.
[17] F. Dammak, L Baccour and A. M. Alimi, A comparative analysis for multi-attribute decision making methods: topsis, ahp, vikor using intuitionistic fuzzy sets, in: *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, 2015.
[18] F. Dammak, L. Baccour and A. M. Alimi, The impact of criterion weights techniques in topsis method of multi-criteria decision making in crisp and intuitionistic fuzzy domains, in: *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, 2015.
[19] L. Dengfeng and C. Chuntian, New similarity measure of intuitionistic fuzzy sets and application to pattern recognitions, *Pattern Recognit. Lett.* **23** (2002), 221–225.
[20] L. Fan and X. Zhangian, Similarity measures between vague sets, *J. Softw.* **6** (2001), 922–927.
[21] H. Freeman, Computer processing of line-drawing images, *ACM Comput. Surv.* **6** (1974), 57–97.
[22] D. H. Hong and C. Kim, A note on similarity measures between vague sets and between elements, *Inform. Sci.* **115** (1999), 83–96.
[23] G. S. Huang, Y.-S. Liu and X.-D. Wang, Some new distances between intuitionistc fuzzy sets, in: *Proc. of Fourth International Conference on Machine Learning and Cybernetics*, pp. 18–21, Guangzhou, 2005.
[24] W.-L. Hung and M.-S. Yang, Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance, *Pattern Recognit. Lett.* **25** (2004), 1603–1611.
[25] W.-L. Hung and M.-S. Yang, Similarity measures of intuitionistic fuzzy sets based on l_p metric, *Approx. Reason.* **46** (2007), 120–136.
[26] W.-L. Hung and M.-S. Yang, On similarity measures between intuitionistic fuzzy sets, *Int. J. Intell. Syst.* **23** (2008), 364–383.
[27] V. Khatibi and G. A. Montazer, Intuitionistic fuzzy set vs. fuzzy set application in medical pattern recognition, *Artif. Intell. Med.* **47** (2009), 43–52.

[28] Z. Liang and P. Shi, Similarity measures on intuitionistic fuzzy sets, *Pattern Recognit. Lett.* **24** (2003), 2687–2693.
[29] Y. Li, D. L. Olson and Z. Qin, Similarity measures between intuitionistc fuzzy (vague) sets: a comparative analysis, *Pattern Recognit. Lett.* **28** (2007), 278–285.
[30] W. Maghrebi, L. Baccour, M. A. Khabou and A. M. Alimi, An indexing and retrieval system of historic art images based on fuzzy shape similarity, in *Proc. 6th Mexican International Conference on Artificial Intelligence, MICAI '2007*, pp. 623–633, 2007.
[31] W. Maghrebi, A. Borchani, M. A. Khabou and A. M. Alimi, A system for historic document image indexing and retrieval based on xml database conforming to mpeg7 standard, in: *GREC 2007, LNCS 5046*, pp. 114–125, 2008.
[32] H. B. Mitchell, On the Dengfeng-Chuntian similarity measure and its application to pattern recognition, *Pattern Recognit. Lett.* **24**, (2003), 3101–3104.
[33] S. Mukherjee and K. Basu, Solution of a class of intuitionistic fuzzy assignment problem by using similarity measures, *Knowl. Based Syst.* **27** (2012), 170–179.
[34] C. P. Pappis and I. Karacapilidis, A comparative assessment of measures of similarity of fuzzy values, *Fuzzy Sets Syst.* **56** (1993), 171–174.
[35] J. H. Park, J. H. Hwang, W. J. Park, H. Wei and S. H. Lee, Similarity measure on intuitionistic fuzzy sets, *J. Cent. South Univ.* **20** (2013), 2233–2238.
[36] M. Pechwitz, S. Snoussi Maddouri, V. Maergner, N. Ellouze and H. Amiri, Ifn/enit – data set of handwritten Arabic words, in: *CIFED 2002*, pp. 129–136, 2002.
[37] S. K. De, R. Biswas and A. R. Roy, An application of intuitionistic fuzzy sets in medical diagnosis, *Fuzzy Sets Syst.* **117** (2001), 209–213.
[38] E. Szmidt and J. Kacprzyk, Intuitionistic fuzzy sets in group decision making, *Notes on IFS* **2** (1996), 11–14.
[39] E. Szmidt and J. Kacprzyk, Distances between intuitionistic fuzzy sets, *Fuzzy Sets Syst.* **114** (2000), 505–518.
[40] E. Szmidt and J. Kacprzyk. Analysis of similarity measures for Atanassov's intuitionistic fuzzy sets. In *IFSA-EUSFLAT 2009*, 2009.
[41] I. K. Vlachos and G. D. Sergiadis, Intuitionistic fuzzy information – applications to pattern recognition, *Pattern Recognit. Lett.* **28** (2007), 197–206.
[42] W. Wang and X. Xin, Distance measure between intuitionistic fuzzy sets, *Pattern Recognit. Lett.* **26** (2005), 2063–2069.
[43] J. Ye, Cosine similarity measures for intuitionistic fuzzy sets and their applications, *Math. Comput. Model.* **53** (2011), 91–97.
[44] C. Zhang and H. Fu, Similarity measures on three kinds of fuzzy sets, *Pattern Recognit. Lett.* **27** (2006), 1307–1317.