sciendo
Control and Cybernetics
vol. 51 (2022) No. 2
pages: 249-266
DOI: 10.2478/candc-2022-0016

Similarity measures for Atanassov's intuitionistic fuzzy sets: some dilemmas and challenges*

by

Eulalia Szmidt[1,2], **Janusz Kacprzyk**[1,2] **and Paweł Bujnowski**[1]

[1]Systems Research Institute Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
[2]Warsaw School of Information Technology, ul. Newelska 6, 01-447 Warsaw, Poland
E-mail: {szmidt, kacprzyk, pbujno}@ibspan.waw.pl

Abstract: We discuss some aspects of similarity measures in the context of Atanassov's intuitionistic fuzzy sets (IFSs, for short). IFSs, proposed in 1983, are a relatively new tool for the modeling and simulation and, because of their construction, present us with new challenges as far the similarity measures are concerned. Specifically, we claim that the distances alone are not a proper measure of similarity for the IFSs. We stress the role of a lack of knowledge concerning elements (options, decisions, etc.) and point out the role of the opposing (complementing) elements. We also pay attention to the fact that it is not justified to talk about similarity when one has not enough knowledge about the compared objects/elements. Some novel measures of similarity are presented.

Keywords: similarity measures, Atanassov's intuitionistic fuzzy sets, distances, complements

1. Introduction

To propose and use a proper similarity measure is both a complex and important task. Much depends on the problem discussed. The similarity measures have been a subject of interest in science for many years and it has been well known that there is no one and only similarity measure. The roots of the notion of similarity are found in the works of Pythagorean philosophers (Reeves, 2020). Since then, a whole array of similarity measures have been proposed, discussed and compared.

In this paper we deal with one, yet very popular type of similarity measures, seen as dual measures of distances. However, this point of view is challenged nowadays. We discuss here measures of similarity where distances are used, but we emphasize

*Submitted: April 2022; Accepted: April 2022

that the complements of the objects (elements, options, etc.) play an important role as well for the very meaning of similarity. Moreover, we show that when we are faced with elements which are difficult to classify (which means that they are from the border regions), it is also difficult to speak about their similarity to other elements. The considerations are presented in the context of Atanassov's intuitionistic fuzzy sets (IFSs, for short). The motivation is that the IFSs are one of significant and widely used extensions of fuzzy sets. They have attracted a lot of attention, this fact being confirmed by many citations. The IFSs are a very convenient tool while making decisions, analyzing data, etc. Their structure renders a way of thinking by a human considering pros, cons, and a lack of knowledge when faced with real problems. A lack of knowledge is a challenge when looking for a suitable measure of similarity. We discuss the problem in detail.

The paper is structured as follows. In Section 2 we briefly recall the basic information about the IFSs, including a geometrical representation, and some notions, which are used in the further considerations. In Section 3 a typical approach for examining the similarity by using distances is discussed, and the importance of taking into account the complement elements is shown. Several measures are presented and discussed. Next, we consider the role of transitivity, which is important in the context of distances, but should be carefully considered in the context of similarity. We also discuss another issue that constitutes a challenge, being the result of the lack of knowledge, occurring in many real tasks, and is intrinsically linked to the IFSs. Finally, Section 5 concludes the paper with a summary.

2. Brief introduction to intuitionistic fuzzy sets

2.1. The preliminaries and the prevoiuos work

One of the possible generalizations of a fuzzy set in X (Zadeh, 1965), given by

$$A^{'} = \{\langle x, \mu_{A^{'}}(x)\rangle | x \in X\} \tag{1}$$

where $\mu_{A^{'}}(x) \in [0,1]$ is the membership function of the fuzzy set $A^{'}$, is an IFS (Atanassov, 1983, 1999, 2012), denoted A, which is given by

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\} \tag{2}$$

where: $\mu_A : X \to [0,1]$ and $\nu_A : X \to [0,1]$ such that

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \tag{3}$$

and $\mu_A(x)$, $\nu_A(x) \in [0,1]$ denote the degree of membership and the degree of non-membership of $x \in A$, respectively. (See Szmidt and Baldwin, 2006, for assigning memberships and non-memberships for IFSs from data.)

For each IFS in X, we will call

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{4}$$

an *intuitionistic fuzzy index* or a *hesitation margin* of $x \in A$, this quantity expressing the lack of knowledge of whether x belongs to A or not (cf. Atanassov, 1999). It is obvious that $0 \leq \pi_A(x) \leq 1$, for each $x \in X$.

The hesitation margin has been shown to be important while considering the distances (Szmidt and Kacprzyk, 1997, 2000, 2005, 2006), entropy (Szmidt and Kacprzyk, 2001, 2007), similarity (Szmidt and Kacprzyk, 2004, 2007b) for the IFSs, etc. i.e., the measures that play a crucial role in virtually all information processing tasks (Szmidt, 2014, and Szmidt and Kacprzyk, 2015).

The hesitation margin turns out to be relevant for applications – in image processing (cf. Bustince et al., 2006), classification of imbalanced and overlapping classes (cf. Szmidt and Kukier, 2006, 2008a,b), the classification via the intuitionistic fuzzy trees (cf. Bujnowski, Szmidt and Kacprzyk, 2014), selection of the best discriminative attributes (Szmidt, Kacprzyk and Bujnowski, 2021), Pearson correlation coefficient (Szmidt and Kacprzyk, 2010a, 2012; Szmidt, Kacprzyk and Bujnowski, 2011a,b, 2012a), Spearman correlation coefficient (Szmidt and Kacprzyk,2010c), Kendall correlation coefficient (Szmidt and Kacprzyk, 2016b,c), Principal Component Analysis (Szmidt and Kacprzyk, 2012a; Szmidt, Kacprzyk and Bujnowski, 2011a,b, 2012a), ranking procedures (Szmidt and Kacprzyk, 2008a,c, 2009a,b,c, 2010b), text categorization (Szmidt and Kacprzyk, 2008b), group decision making (e.g., Atanassova, 2004), genetic algorithms (see, for instance, Roeva and Michalikova, 2013), negotiations, consensus reaching, voting, etc. It is worth mentioning that the approaches referred to above were successfully applied for benchmark data from the UCI Machine Learning Repository (www.ics.uci.edu/ mlearn/).

Certainly, each fuzzy set may be represented by the following IFS

$$A = \{< x, \mu_{A'}(x), 1 - \mu_{A'}(x) > | x \in X\}. \tag{5}$$

On the other hand, for each fuzzy set $A^{'}$ in X, we evidently have

$$\pi_{A'}(x) = 1 - \mu_{A'}(x) - [1 - \mu_{A'}(x)] = 0 \text{ for each } x \in X. \tag{6}$$

The application of IFSs instead of fuzzy sets means the introduction of another degree of freedom into the description of a set. Such a generalization of fuzzy sets gives us an additional possibility to represent imperfect knowledge, what leads to describing many real problems in a more adequate way.

Basically, the IFSs based models may be adequate in situations when we face human testimonies, opinions, etc. involving answers of three types:

- yes,
- no,
- I do not know, I am not sure, etc.

Voting can be a good example of such a situation, as the human voters may be divided into three groups of those who:

- vote for,
- vote against,

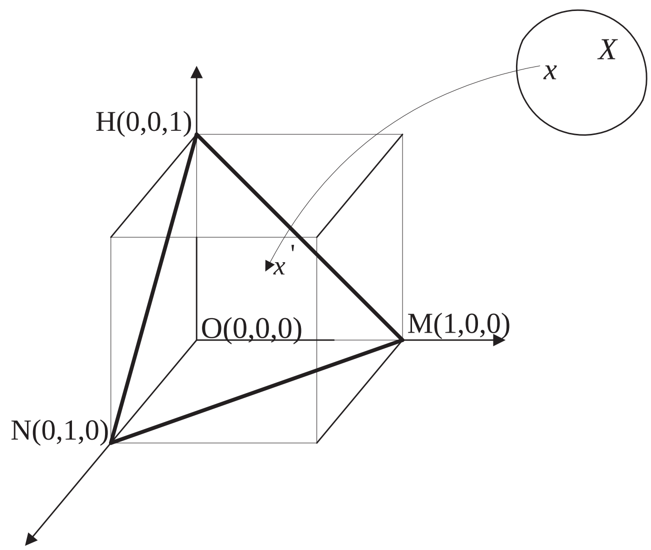

Figure 1. Geometrical representation of IFSs in 3D

- abstain or give invalid votes.

As noted already, applications of IFSs to group decision making, negotiations, and other real situations are presented in Szmidt and Kacprzyk's papers, listed in the references.

2.2. Geometrical representation of IFSs

Since for each element x, belonging to an IFS, the values of membership $\mu(x)$, non-membership $\nu(x)$, and the intuitionistic fuzzy index $\pi(x)$ sum up to one, i.e.

$$\mu(x) + \nu(x) + \pi(x) = 1 \tag{7}$$

and $\mu_x(x)$, $\nu_x(x)$, $\pi(x) \in [0, 1]$ we can imagine a unit cube (Fig. 1). Inside the cube there is an MNH triangle, where equation (7) is fulfilled (Szmidt and Kacprzyk, 2000, Szmidt, 2014). Consequently, the MNH triangle represents the surface, within which coordinates of any element belonging to an IFS can be represented. Each point belonging to the MNH triangle is described via three coordinates: (μ, ν, π). Point $M(1, 0, 0)$ represents the elements fully belonging to an IFS as $\mu = 1$. Point $N(0, 1, 0)$ represents the elements fully not belonging to an IFS as $\nu = 1$. Point $H(0, 0, 1)$ represents the elements about which we are not able at all to say if they belong or not to an IFS (intuitionistic fuzzy index $\pi = 1$). The segment MN (where $\pi = 0$) represents the elements belonging to the classical fuzzy sets ($\mu + \nu = 1$).

Any combination of the values characterizing an IFS can be represented inside the triangle MNH. This means that each element belonging to an IFS can be represented as a point (μ, ν, π) belonging to the interior of the triangle MNH (Figure 1).

REMARK 1 *We use the capital letters (e.g., M, N, H) for the geometrical representation of x_i's (Fig. 1) on the plane. The same notation (capital letters) is used in the paper for sets, but we always explain the current meaning of a symbol used.*

In our further considerations we will use the notion of distance and of the complement of an element.

Following the three term description of the IFSs (cf. Szmidt 2014, Szmidt and Kacprzyk, 2000, 2010c, 2011a; Szmidt, Kacpszyk and Bujnowski, 2020) the Hamming distance between IFSs A and B is

$$l^1_{IFS}(A,B) = \frac{1}{2n}\sum_{x\in E} \mid \mu_A(x) - \mu_B(x) \mid + \mid \nu_A(x) - \nu_B(x) \mid + \mid \pi_A(x) - \pi_B(x) \mid; \tag{8}$$

the Euclidean distance is

$$q^1_{IFS}(A,B) =$$

$$\sqrt{\frac{1}{2n}(\sum_{x\in E}(\mu_A(x)-\mu_B(x))^2 + (\nu_A(x)-\nu_B(x))^2) + (\pi_A(x)-\pi_B(x))^2)}; \tag{9}$$

the Hausdorff distance, for the normalized Hamming distance expressed in the spirit of (Szmidt and Kacprzyk, 2000, 2006), given by (8) (see Szmidt and Kacprzyk, 2011a), is

$$\begin{aligned} H_3(A,B) &= \frac{1}{n}\sum_{i=1}^{n} \max\left\{|\mu_A(x_i)-\mu_B(x_i)|, |\nu_A(x_i)-\nu_B(x_i)|,\right. \\ &\quad \left.|\pi_A(x_i)-\pi_B(x_i)|\right\} \end{aligned} \tag{10}$$

where $l^1_{IFS}(A,B)$, $q^1_{IFS}(A,B)$, and $H_3(A,B) \in [0,1]$ and fulfill all the properties of distances. In the above formulae n is the number of elements.

The complement A^C of an IFS A is

$$A^C = \{\langle x, \nu_A(x), \mu_A(x), \pi_A(x)\rangle | x \in X\}. \tag{11}$$

Accounting for the complement elements in the similarity measures seems important in many tasks. For example, in image recognition, the most "dissimilar" image is a negative image which can be understood as an image consisting of the complement elements.

3. A typical approach via measuring a distance

A similarity measure is often defined via a distance measure, and – intuitively – the smaller the distance, the bigger the similarity. However, to be a proper measure of similarity, such a measure is not always required to satisfy all the distance axioms. The similarity has typically been assumed to be symmetric. Tversky (1977), however, has provided some empirical evidence that similarity should not always be treated as a symmetric relation. For example, we may say that Betty is similar to her mother but we do not say that Betty's mother is similar to Betty.

Besides symmetry, also transitivity is not always welcome. A well-known example is that human beings and horses are different. However, both are similar to centaurs. So a proper distance reflecting the similarity should be small from the humans to centaurs, and from the horses to centaurs, but large from the humans to horses.

We recall this to show that a similarity measure may have some features, which can be useful in some situations, but are not welcome in other cases (see Cross and Sudkamp, 2002; Wang, De Baets and Kerre, 1995; Veltkamp, 2001a,b, and Veltkamp and Hagedoorn, 2000).

Now we will present an example showing that a distance alone is not the best measure of similarity.

EXAMPLE 1 Suppose we wish to compare three items described as $x_i(\mu_i, \nu_i, \pi_i)$, $i = 1, 2, 3$, to find out if item x_1 is more similar to item x_2 or to item x_3. Consider, for simplicity, one attribute only (or a properly aggregated set of attributes) so that the particular items are described as

- item $x_1(0.2, 0.2, 0.6)$
- item $x_2(0.3, 0.4, 0.3)$
- item $x_3(0.1, 0.6, 0.3)$.

The respective distances between the attributes describing the above items are equal to, in terms of (8):

$$l^1_{IFS}(x_1, x_2) = \frac{1}{2}(|0.2 - 0.3| + |0.2 - 0.4| + |0.6 - 0.3|) = 0.3 \quad (12)$$

$$l^1_{IFS}(x_1, x_3) = \frac{1}{2}(|0.2 - 0.1| + |0.2 - 0.6| + |0.6 - 0.3|) = 0.4. \quad (13)$$

As $l^1_{IFS}(x_1, x_2)$ is smaller than $l^1_{IFS}(x_1, x_3)$, we can come to the conclusion (using the distances) that the items x_1 and x_2 are more similar than x_1 and x_3.

However, when saying something about the similarity of the elements we should have in mind similarity regarding their complements, too. If a distance between element x and another element y is the same as the distance between element x and the complement of y, we can not speak about similarity between x and y. It is justified that an element y and its complement y^C should be classified to different classes. Hence, we are not able to classify x if its similarities to y and to y^C are the same.

Having the above in mind we verify the distances between x_1 and the complements of x_2 and x_3, where:

$$x_2^C = (0.4, 0.3, 0.3)$$

and

$$x_3^C = (0.6, 0.1, 0.3).$$

As a result we obtain, from (8):

$$l^1_{IFS}(x_1, x_2^C) = 0.3 \tag{14}$$

$$l^1_{IFS}(x_1, x_3^C) = 0.4 \tag{15}$$

which means that the distance between x_1 and x_2, (12), is equal to the distance between x_1 and x_2^C, (14). The same situation occurs in the case of x_1 and x_3, (13), for which the distance is equal to the distance between x_2 and x_3^C, (15). It is difficult to agree that in such a situation we can speak about a high similarity of the items.

The above examples justify the following conclusions:

- if a distance between two (or more) elements, or objects, is big, then the similarity does not exist, i.e. it is too small to be treated as proper "similarity";
- if a distance is small, it is difficult to determine the similarity having in mind a pure distance only; the distance can be small and the compared objects can be more dissimilar than similar.

The above considerations point out that a properly constructed similarity measure should take into account in addtion to the distance between the objects, also the distance to their complements. The measure of similarity between the IFSs, presented by Szmidt and Kacprzyk (2004a,b) follows this intuitively appealing rule.

Let us calculate the similarity of any two elements belonging to an IFS, which are geometrically represented by points X and F (Fig. 2) belonging to the triangle MNH. The proposed measures indicate whether X is more similar to F or to F^C, where F^C is the complement of F. In other words, the proposed measures answer the question whether X is more similar or more dissimilar to F (Fig. 2), expressed as:

$$Sim_{rule}(X, F) = \frac{l^1_{IFS}(X, F)}{l^1_{IFS}(X, F^C)} \tag{16}$$

where: $l^1_{IFS}(X, F)$ is a distance from $X(\mu_X, \nu_X, \pi_X)$ to $F(\mu_F, \nu_F, \pi_F)$,

$l^1_{IFS}(X, F^C)$ is a distance from $X(\mu_X, \nu_X, \pi_X)$ to $F^C(\nu_F, \mu_F, \pi_F)$,

F^C, (11), is a complement of F, distances $l^1_{IFS}(X, F)$ and $l^1_{IFS}(X, F^C)$ being calculated from (8).

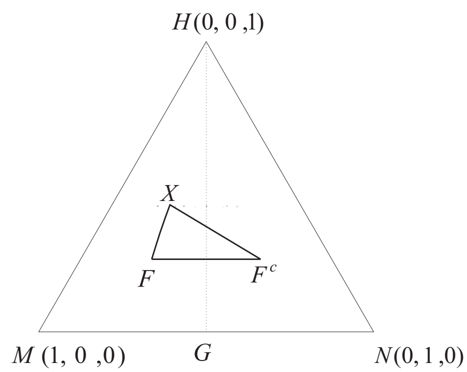

Figure 2. The triangle MNH explaining the ratio-based measure of similarity

The following conditions are fulfilled for (16)

$$0 \leq Sim_{rule}(X, F) \leq \infty \tag{17}$$

$$Sim_{rule}(X, F) = Sim_{rule}(F, X). \tag{18}$$

Szmidt and Kacprzyk (2004a) have noticed that the formula (16) can also be stated as

$$\begin{aligned} Sim_{rule}(X, F) &= \frac{l^1_{IFS}(X, F)}{l^1_{IFS}(X, F^C)} = \frac{l^1_{IFS}(X^C, F^C)}{l^1_{IFS}(X, F^C)} = \\ &= \frac{l^1_{IFS}(X, F)}{l^1_{IFS}(X^C, F)} = \frac{l^1_{IFS}(X^C, F^C)}{l^1_{IFS}(X^C, F)}. \end{aligned} \tag{19}$$

Certainly, we assume that the denominators in (19) are not equal to zero.

It can be noticed that

- if X and F are identical, then $Sim_{rule}(X, F) = 0$;
- if X is to the same extent similar to F and F^C (the respective distances are equal), then $Sim_{rule}(X, F) = 1$;
- if X and F^C are closer than X and F, then $Sim_{rule}(X, F) > 1$;
- if $X = F^C$ (or $X^C = F$), then there is $l^1_{IFS}(X, F^C) = l^1_{IFS}(X^C, F) = 0$ which means the complete dissimilarity of X and F (or in other words, the identity of X and F^C), and then $Sim_{rule}(X, F) \to \infty$;
- $X = F = F^C$ means the highest possible entropy (see (16)) for both elements F and X i.e. the highest “fuzziness” – not too constructive a case when looking for both the similarity and dissimilarity.

From the above properties it follows that while applying the measure (16) to the analysis of the similarity of two objects, one should be interested in the values

$$0 \leq Sim_{rule}(X, F) < 1$$

where 0 means the highest similarity, and the values close to 1 mean a very low similarity.

The proposed measure (16) has been constructed for selecting the objects, which are more similar than dissimilar [and well-defined in the sense of possessing (or not) attributes we are interested in].

Now, returning to Example 1, we will show that a measure of similarity, defined as above, i.e. through (16), between two elements belonging to an IFS, is more powerful than a simple distance between them.

Using the data from Example 1 we obtain from (16) the following:

$$Sim(x_1, x_2) = l^1_{IFS}(x_1, x_2)/l^1_{IFS}(x_1, x_2^C) = 1 \tag{20}$$

$$Sim(x_1, x_3) = l^1_{IFS}(x_1, x_3)/l^1_{IFS}(x_1, x_3^C) = 1 \tag{21}$$

which means that in both cases the similarity is the same and very weak, despite the small distances between the compared items. The similarity between x_1 and x_2 is very weak as x_1 is similar to the same extent to x_2 and x_2^C. The same concerns x_1 and x_3.

4. Other similarity measures including the concept of a complement

4.1. The definitions and the measures

The similarity measure (16) properly reflects our intuition concerning the similarity, but it does not follow the range of the usually assumed values for the similarity measures.

To be consistent with the common scientific tradition (i.e. using a similarity measure(s) whose numerical values belong to $[0, 1]$), and at the same time preserve the advantages of the measure (16), we were looking for a function using the same two kinds of distances as in (16) (i.e., $l^1_{IFS}(X, F)$, $l^1_{IFS}(X, F^C)$), but with values of the measure from the interval $[0, 1]$. Specifically, following Szmidt and Kacprzyk (2007b), we have

$$f(l^1_{IFS}(X, F), l^1_{IFS}(X, F^C)) = \frac{l^1_{IFS}(X, F)}{l^1_{IFS}(X, F) + l^1_{IFS}(X, F^C)}. \tag{22}$$

The above function, (22), is constructed under the condition that the case when $X = F = F^C$ (which is, by obvious reasons, not interesting in practice) is excluded from the considerations. The assumption $X = F = F^C$ means that one tries to compare an element (represented by) X, which it is impossible to classify, as F and F^C should belong to different classes. $F = F^C$ in terms of geometrical representation in Fig. 2 means that X, F and F^C, representing respective elements from the IFS are at the same point on the HG segment. So the cases for which $l^1_{IFS}(X, F) = l^1_{IFS}(X, F^C) = 0$ are excluded from considerations.

Returning to (22), the function takes into account the same two distances as the previous measure, (16), but now the new measure is normalized (its values are in the interval $[0,1]$) (see Szmidt and Kacprzyk, 2007b). It is obvious that (22) is a concept, which is dual to the similarity measure. If (22) is equal to zero, then the similarity is equal to 1; if (22) is equal to 1, then the similarity is equal to zero. In other words, we may use (22) to construct a similarity measure. Since

$$0 \leq f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C)) \leq 1, \tag{23}$$

then we look for a monotone decreasing function g, fulfilling:

$$g(1) \leq g(f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))) \leq g(0). \tag{24}$$

and from the above it follows that

$$0 \leq g(f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))) - g(1) \leq g(0) - g(1) \tag{25}$$

$$0 \leq \frac{g(f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))) - g(1)}{g(0) - g(1)} \leq 1. \tag{26}$$

As a result, we get a function having the properties of a similarity measure, i.e., a monotone decreasing function of (22).

DEFINITION 1 (SZMIDT AND KACPRZYK, 2007B)

$$Sim(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C)) = \frac{g(f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))) - g(1)}{g(0) - g(1)} \tag{27}$$

where $f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))$ *is given by (22).*

A simple function g, which may here be applied is

$$g(x) = 1 - x \tag{28}$$

which gives, from (27) (see, again, Szmidt and Kacprzyk, 2007b),

$$\begin{aligned} Sim_1(X,F) &= Sim(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C)) = \\ &= 1 - f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C)) = 1 - \frac{l^1_{IFS}(X,F)}{l^1_{IFS}(X,F) + l^1_{IFS}(X,F^C)}. \end{aligned} \tag{29}$$

Another function $g(x)$ can be defined as

$$g(x) = \frac{1}{1+x}, \tag{30}$$

yielding (Szmidt and Kacprzyk, 2007b)

$$\begin{aligned} Sim_2(X,F) &= Sim(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C)) = \\ &= \frac{1 - f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))}{1 + f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))}. \end{aligned} \tag{31}$$

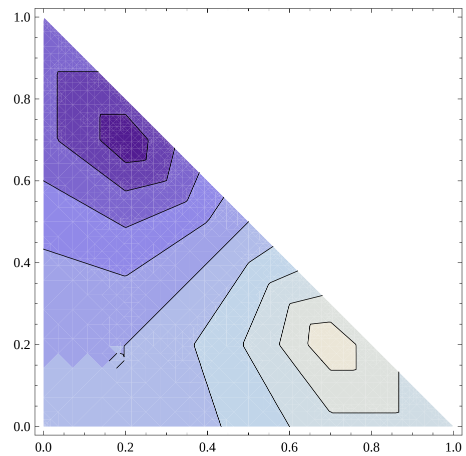

Figure 3. Contourplot of measure (4.1) for any element from an IFS and (0.7, 0.2, 0.1)

Then, the function

$$g(x) = \frac{1}{1+x^2} \tag{32}$$

leads to (see, once again, Szmidt and Kacprzyk, 2007b)

$$\begin{aligned} &Sim_3(X,F) = Sim(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C)) = \\ &= \frac{1 - f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))^2}{1 + f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))^2}. \end{aligned} \tag{33}$$

It is possible to use, as well, $g(x) = \frac{1}{1+x^n}$ where $n = 3, 4, \ldots, k$, but the counterpart similarity measures $(\frac{1-x^n}{1+x^n})$ give the values, which are less convenient for the comparison when the values of x are small (Szmidt and Kacprzyk, 2007b).

The exponential function (cf. Pal and Pal, 1991) is another one which may be applied

$$g(x) = e^{-x}, \tag{34}$$

giving for (22) (Szmidt and Kacprzyk, 2007b)

$$\begin{aligned} Sim_4(X,F) &= Sim(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C)) = \\ &= \frac{e^{-f(l^1_{IFS}(X,F), l^1_{IFS}(X,F^C))} - e^{-1}}{1 - e^{-1}}. \end{aligned} \tag{35}$$

Certainly, one could continue generating more complicated functions $g(x)$ (being the decreasing functions of f), but it would not give any additional insight as far as the similarity is concerned.

The similarity measures (29) – (35) satisfy the following properties:

$$Sim_i(X,F) \in [0,1] \tag{36}$$
$$Sim_i(X,X) = 1 \tag{37}$$
$$Sim_i(X,X^C) = 0 \tag{38}$$
$$Sim_i(X,F) = Sim_i(F,X) \tag{39}$$

for $i = 1, \dots, 4$.

The similarity measures, discussed in this section, evaluate the similarity of any two elements (X and F) belonging to an IFS. The corresponding similarity measures for the IFSs A and B, containing n elements each, are given by the following formula:

$$Sim_k(A,B) = \frac{1}{n}\sum_{i=1}^{n} Sim_k(l_{IFS}^1(X_i,F_i), l_{IFS}^1(X_i,F_i^C)) \tag{40}$$

for $k = 1, \dots, 4$.

Although in the formulas presented above we use the normalized Hamming distance, it is possible to replace it by other kinds of distances, too.

To be more specific, the function $f(l_{IFS}^1(X,F), l_{IFS}^1(X,F^C))$, given by (22), with the Hamming distance used in (29) – (35), can be replaced by the corresponding function with the Euclidean distance, i.e.:

$$f(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C)) = \frac{q_{IFS}^1(X,F)}{q_{IFS}^1(X,F) + q_{IFS}^1(X,F^C)} \tag{41}$$

where $q_{IFS}^1(X,F)$ is given by (9). For example, the measure corresponding to the similarity measure (29), in which (41) instead of (22) is applied, is:

$$\begin{aligned} Sim_1(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C)) &= 1 - f(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C)) \\ &= 1 - \frac{q_{IFS}^1(X,F)}{q_{IFS}^1(X,F) + q_{IFS}^1(X,F^C)}. \end{aligned} \tag{42}$$

The measure, corresponding to the similarity measure (31), in which (41) instead of (22) is applied, is:

$$Sim_2(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C)) = \frac{1 - f(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C))}{1 + f(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C))}. \tag{43}$$

The measure, corresponding to the similarity measure (33), in which (41) instead of (22) is applied, is:

$$Sim_3(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C)) = \frac{1 - f(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C))^2}{1 + f(q_{IFS}^1(X,F), q_{IFS}^1(X,F^C))^2}. \tag{44}$$

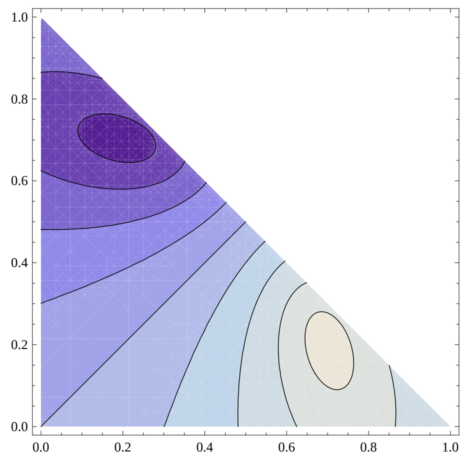

Figure 4. Contourplot of (4.1) for any element from an IFS and (0.7, 0.2, 0.1)

The measure, corresponding to the similarity measure (33), in which (41) instead of (22) is applied, is:

$$Sim_4(q^1_{IFS}(X,F), q^1_{IFS}(X,F^C)) = \frac{e^{-f(q^1_{IFS}(X,F), q^1_{IFS}(X,F^C))} - e^{-1}}{1-e^{-1}}. \quad (45)$$

We can also introduce other measures of similarity using the Hausdorff distance (cf. Grünbaum, 1967). We have shown (see Szmidt and Kacprzyk, 2011a) that in the case of the Hausdorff distance between the IFSs we should use a formula with all three terms describing the sets. If in the formulas (29) – (35) we replace (22) by (10), the new similarity measures, referring to the Hausdorff distance, are obtained. For example, the counterpart of (29) with (10) replacing (22) is:

$$\begin{aligned} Sim(H_3(X,F), H_3(X,F^C)) &= 1 - f(H_3(X,F), H_3(X,F^C)) = \\ &= 1 - \frac{H_3(X,F)}{H_3(X,F) + H_3(X,F^C)}. \end{aligned} \quad (46)$$

In Fig. 5 we show an example of results, implied by (46) – the presence of the complement of an element and its influence on the results are visible.

4.2. The transitivity and the lack of knowledge

The similarity measures discussed here (that is, in Section 4), take into account not only the relation to an element we are interested in but also that to its complement. As a result, the measures discussed here meet better our expectations than the similarity measures that are just dual to the distance (cf. Example 1). For example, we avoid high values of the similarity of an element and its complement. However, we should still use the similarity measures carefully.

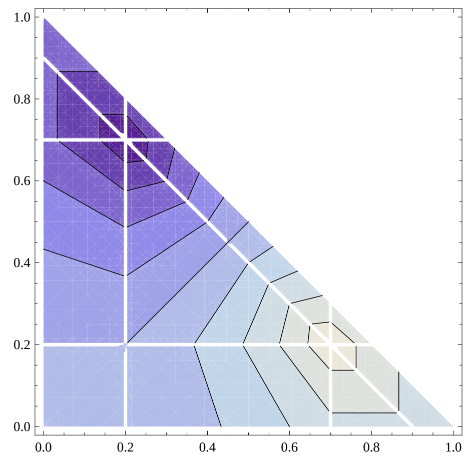

Figure 5. Contourplot of (46) for any element from an IFS and (0.7, 0.2, 0.1)

The question arises what should be done if we wish, e.g., to use the similarity measure (42) and to differentiate between the elements $(0.3, 0, 0.7)$ and $(0.5, 0.4, 0.1)$, which are obviously different from the point of view of decision making, but both are similar to the element $(0.7, 0.2, 0.1)$ to the same extent, equal to 0.6 (cf. Fig. 4). Most important is that we should not determine the similarity of $(0.3, 0, 0.7)$ and $(0.5, 0.4, 0.1)$ before calculating their direct similarity from (42), whereupon we obtain the value of 0.51 (different from 0.6). This observation about examining similarity seems important when one tries to conclude about the similarity of different elements based on their direct distances to "the ideal" element (1,0,0). The transitivity is not always justified and should not be automatically applied. A good example is provided by the existing similarities of: humans to centaurs and horses to centaurs, while the similarity between humans and horses does not exist.

Another important issue is the lack of knowledge.

The IFSs are a specific tool for modeling, making it possible to represent different levels of lack of knowledge. From full knowledge, concerning an element (which can represent an option, alternative, etc.), to a complete lack of knowledge. In this context, it is important not to "mechanically" treat the notion of similarity. For example, if we have two elements about which we know nothing, i.e., $x_1(0, 0, 1)$ and $x_2(0, 0, 1)$, then we could say that they are formally similar ("the same"). However, the elements can be the same or completely different as we know nothing about them, and hence we cannot compare them.

The above considerations concern, in some sense, other situations, too. What can we say about, e.g., $x_1(0.4, 0.4, 0.2)$, $x_2(0.42, 0.38, 0.2)$, and $x_3(0.38, 0.42, 0.2)$? The distances between x_1 and x_2, and x_1 and x_3 are small, so that we could say that all the elements could be considered very similar. However, x_2 is closer to (1, 0, 0), whereas

x_3 is closer to (0, 1, 0). The conclusion is that if μ is close to ν, it is difficult to consider similarity (here x_i is close to x_i^C for $i = 1, 2, 3$).

The similarity measures are just a basis for the classification tasks. As it is difficult to classify the elements close to the segment GH in Fig. 2, then we should not speak about the similarity of the elements belonging to this region (for which μ is close to ν). Moreover, for the elements with high values of the hesitation margins π, we should also be careful when speaking about similarity. For example, if $x(\mu, \nu, \pi) = (0.3, 0.1, 06)$, then we can be faced with the situation when μ could become $\mu+\pi$ giving $(0.9, 0.1)$ or ν could become $\nu + \pi$ giving $(0.3, 0.7)$ or, e.g., μ could become $\mu + 0.5\pi$ and ν could become $\nu + 0.5\pi$ giving $(0.6, 0.4)$. All the three possibilities: $(0.9, 0.1)$, $(0.3, 0.7)$, and $(0.6, 0.4)$ are different, and we do not know which one might be real (in fact we could consider an infinite number of other possibilities depending on how the hesitation margin could be divided between μ and ν).

5. Conclusions

We have discussed some selected measures of similarity for the IFSs. It is important to emphasize that a concrete "tool" has been considered, namely the IFSs. The possibilities and properties of a tool imply, in a sense, which measures of similarity should be used. Certainly, a purpose is very important, too. We considered the measures of similarity using distances, but not in a standard way, that is when the similarity measure is a dual measure to the distance (the greater the distance, the smaller the similarity). We have emphasized, as well, the importance of the complements in the IFS similarity measures, which make use of the distances. We have also pointed out that for some elements, belonging to an IFS, looking for similar elements is in some sense an ill-defined problem.

References

ATANASSOV, K. (1983) Intuitionistic Fuzzy Sets. VII ITKR Session. Sofia *(Centr. Sci.-Techn. Libr. of Bulg. Acad. of Sci.*, 1697/84) (in Bulgarian).

ATANASSOV, K. T. (1999) *Intuitionistic Fuzzy Sets: Theory and Applications*. Springer-Verlag.

ATANASSOV, K.T. (2012) *On Intuitionistic Fuzzy Sets Theory*. Springer.

ATANASSOVA, V. (2004) Strategies for Decision Making in the Conditions of Intuitionistic Fuzziness. *Int. Conf. 8th Fuzzy Days*, Dortmund, Germany, 263–269.

BUJNOWSKI, P., SZMIDT, E. AND KACPRZYK, J. (2014) Intuitionistic Fuzzy Decision Trees - a new Approach. In: L. Rutkowski et al., eds.: *Artificial Intelligence and Soft Computing, Part I*. Springer, Switzerland, 181–192.

BUSTINCE, H., MOHEDANO, V., BARRENECHEA, E., PAGOLA, M. (2006) An algorithm for calculating the threshold of an image representing uncertainty through A-IFSs. *IPMU'2006*, 2383–2390.

CROSS, V. AND SUDKAMP, T. (2002) *Similarity and Compatibility in Fuzzy Set Theory*. Physica-Verlag.

GRÜNBAUM, B. (1967) *Convex Polytopes*. Wiley-Interscience, New York.

PAL, N.R. AND PAL, S.K. (1991) Entropy: a new definition and its applications. *IEEE Trans. on Systems, Man, and Cybernetics*, **21**, 5, 1260–1270.

REEVES, J. (2020) *The Science and Religion Dialogue as Natural Philosophy*. Metanexus, https://www.metanexus.net/science-and-religion-dialogue-naturalphilosophy/.

ROEVA, O. AND MICHALIKOVA, A. (2013) Generalized net model of intuitionistic fuzzy logic control of genetic algorithm parameters. In: *Notes on IFSs*, **19** (2), 71–76.

SZMIDT, E. (2014) *Distances and Similarities in Intuitionistic Fuzzy Sets*. Springer.

SZMIDT, E. AND BALDWIN, J. (2006) Intuitionistic Fuzzy Set Functions, Mass Assignment Theory, Possibility Theory and Histograms. *2006 IEEE World Congress on Computational Intelligence*, 237–243.

SZMIDT, E. AND KACPRZYK, J. (1997) On measuring distances between intuitionistic fuzzy sets. *Notes on IFS*, **3**(4), 1–13.

SZMIDT, E. AND KACPRZYK, J. (2000) Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, **114**(3), 505–518.

SZMIDT, E. AND KACPRZYK, J. (2001) Entropy for intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, **118** (3), 467–477.

SZMIDT, E. AND KACPRZYK, J. (2004a) Similarity of intuitionistic fuzzy sets and the Jaccard coefficient. *IPMU 2004*, 1405–1412.

SZMIDT, E. AND KACPRZYK, J. (2004b) A Concept of Similarity for Intuitionistic Fuzzy Sets and its use in Group Decision Making. *2004 IEEE Conf. on Fuzzy Systems*, Budapest, 1129–1134.

SZMIDT, E. AND KACPRZYK, J. (2005) Distances Between Intuitionistic Fuzzy Sets and their Applications in Reasoning In: S. K. Halgamuge and L. Wang, eds., *Computational Intelligence for Modelling and Prediction*, Springer, 101–116.

SZMIDT, E. AND KACPRZYK, J. (2006) Distances Between Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. *IEEE IS'06*, 716–721.

Szmidt, E. and Kacprzyk, J. (2007a) Some problems with entropy measures for the Atanassov intuitionistic fuzzy sets. *Applications of Fuzzy Sets Theory*. *LNAI 4578*, Springer-Verlag, 291–297.

SZMIDT, E. AND KACPRZYK, J. (2007b) A New Similarity Measure for Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. *2007 IEEE Conf. on Fuzzy Systems*, 481–486.

SZMIDT, E. AND KACPRZYK, J. (2008a) Ranking alternatives expressed via intuitionistic fuzzy sets. *12th International Conference IPMU 2008*, 1604–1611.

SZMIDT, E. AND KACPRZYK, J. (2008b) Using intuitionistic fuzzy sets in text categorization In: L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh and J. M. Zurada (eds) *Artificial Intelligence and Soft Computing – ICAISC 2008. Lecture Notes in Computer Science*, **5097**, Springer, Berlin, Heidelberg, 351–362.

SZMIDT, E. AND KACPRZYK, J. (2009a) A method for ranking alternatives expressed via Atanassov's intuitionistic fuzzy sets. In: K. T. Atanassov, O. Hryniewicz, J. Kacprzyk, M. Krawczak, Z. Nahorski, E. Szmidt and S. Zadrozny (Eds.): *Advances in Fuzzy Sets, Intuitionistics Fuzzy Sets, Generalized Nets and Related*

Topics. Academic Publishing House EXIT, Warsaw 2009. Series: Challenging Problems of Science - Computer Science, 161–173.

SZMIDT, E. AND KACPRZYK, J. (2008c) A new approach to ranking alternatives expressed via intuitionistic fuzzy sets. In: D. Ruan et al. (Eds.) *Computational Intelligence in Decision and Control*. World Scientific, 265–270.

SZMIDT, E. AND KACPRZYK, J. (2009b) Ranking of Intuitionistic Fuzzy Alternatives in a Multi-criteria Decision Making Problem. *NAFIPS 2009*, Cincinnati, USA, IEEE, ISBN: 978-1-4244-4577-6.

SZMIDT, E. AND KACPRZYK, J. (2009c) Amount of information and its reliability in the ranking of Atanassov's intuitionistic fuzzy alternatives. In: E. Rakus-Andersson, R. Yager, N. Ichalkaranje and L. C. Jain (Eds.) *Recent Advances in decision Making*, **SCI 222**. Springer-Verlag, 7–19.

SZMIDT, E. AND KACPRZYK, J. (2010a) Correlation between intuitionistic fuzzy sets. In: E. Hullermeier, R. Kruse and F. Hoffmann, eds., LNAI **6178** *(Computational Intelligence for Knowledge-Based Systems Design)*, 169–177.

SZMIDT, E. AND KACPRZYK, J. (2010b) On an Enhanced Method for a More Meaningful Ranking of Intuitionistic Fuzzy Alternatives. *Lecture Notes in Artificial Intelligence*, **6113**, Springer, 232–239.

SZMIDT, E. AND KACPRZYK, J. (2010c) The Spearman rank correlation coefficient between intuitionistic fuzzy sets. In: *Proc. 2010 IEEE Int. Conf. on Intelligent Systems IEEE'IS 2010*, London, 276–280.

SZMIDT, E. AND KACPRZYK, J. (2011a) Intuitionistic fuzzy sets – Two and three term representations in the context of a Hausdorff distance. *Acta Universitatis Matthiae Belii, Series Mathematics*, **19**, 19, 53–62. https://actamath.savbb.sk/pdf/acta1908.pdf

Szmidt, E. and Kacprzyk, J. (2011b) The Kendall Rank Correlation between Intuitionistic Fuzzy Sets. In: *Proc.: World Conference on Soft Computing*, San Francisco, CA, USA, 23/05/2011-26/05/2011.

SZMIDT, E. AND KACPRZYK, J. (2011c) The Spearman and Kendall rank correlation coefficients between intuitionistic fuzzy sets. In: *Proc. 7th conf. European Society for Fuzzy Logic and Technology*, Aix-Les-Bains, France, Antantic Press, 521–528.

SZMIDT, E., KACPRZYK, J. AND BUJNOWSKI, P. (2011a) Pearson's coefficient between intuitionistic fuzzy sets. *Notes on Intuitionistic Fuzzy Sets*, **17**(2), 25–34.

SZMIDT, E., KACPRZYK, J. AND BUJNOWSKI, P. (2011b) Pearson's Correlation Coefficient between Intuitionistic Fuzzy Sets: an Extended Theoretical and Numerical Analysis. In: K. T. Atanassov et al. (Eds.) *Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics*. SRI PAS, Warsaw, 223–236.

SZMIDT, E. AND KACPRZYK, J. (2012a) A New Approach to Principal Component Analysis for Intuitionistic Fuzzy Data Sets. In: S. Greco et al. (Eds.), *IPMU 2012, Part II*, **CCIS 298**, 529–538, Springer-Verlag, Berlin Heidelberg.

SZMIDT, E. AND KACPRZYK, J. (2012b) On an Enhanced Method for a More Meaningful Pearson's Correlation Coefficient between Intuitionistic Fuzzy Sets. *ICAISC (1)*, 334–341.

SZMIDT, E. AND KACPRZYK, J. (2015) Two and three term representations of intuitionistic fuzzy sets: Some conceptual and analytic aspects. *IEEE Int. Conf. on Fuzzy Systems FUZZ-IEEE 2015*, 1–8.

SZMIDT, E., KACPRZYK, J. AND BUJNOWSKI, P. (2012a) Correlation between Intuitionistic Fuzzy Sets: Some Conceptual and Numerical Extensions. *WCCI 2012 IEEE World Congress on Computational Intelligence*, Brisbane, Australia, 480–486.

SZMIDT, E., KACPRZYK, J. AND BUJNOWSKI, P. (2012b) Advances in Principal Component Analysis for Intuitionistic Fuzzy Data Sets. *2012 IEEE 6th International Conference "Intelligent Systems"*, 194–199.

SZMIDT, E., KACPRZYK, J. AND BUJNOWSKI, P. (2020) Attribute Selection for Sets of Data Expressed by Intuitionistic Fuzzy Sets. *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–7.

SZMIDT, E., KACPRZYK, J. AND BUJNOWSKI, P. (2021) Three term attribute description of Atanassov's Intuitionistic Fuzzy Sets as a basis of attribute selection. *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6.

SZMIDT, E. AND KUKIER, M. (2006) Classification of Imbalanced and Overlapping Classes using Intuitionistic Fuzzy Sets. *IEEE IS'06*, London, 722–727.

SZMIDT, E. AND KUKIER, M. (2008a) A New Approach to Classification of Imbalanced Classes via Atanassov's Intuitionistic Fuzzy Sets. In: Hsiao-Fan Wang (Ed.), *Intelligent Data Analysis : Developing New Methodologies Through Pattern Discovery and Recovery*. Idea Group, 85–101.

SZMIDT, E. AND KUKIER, M. (2008b) Atanassov's intuitionistic fuzzy sets in classification of imbalanced and overlapping classes. In: P. Chountas, I. Petrounias, J. Kacprzyk (Eds.): *Intelligent Techniques and Tools for Novel System Architectures*. Springer, Berlin Heidelberg, 455–471.

TVERSKY, A. (1977) Features of similarity. *Psychol. Rev.*, 84, 327–352.

VELTKAMP, R.C. AND HAGEDOORN, M. (2000) Shape similarities, properties, and constructions. In: *Advances in Visual Information Systems, Proc. 4th International Conference, VISUAL 2000*, **LNCS 1929**, Springer, 467–476.

VELTKAMP, R. (2001a) Shape Matching: similarity measures and algorithms. *Proc. Shape Modelling International*, Italy, IEEE Press, 187–197.

VELTKAMP, R.C. (2001b) Shape Matching: Similarity Measures and Algorithms. UU-CS-2001-3, 1–17.

WANG, X., DE BAETS, B. AND KERRE, E. (1995) A comparative study of similarity measures. *Fuzzy Sets and Systems*, **73** (2), 259–268.

ZADEH, L.A. (1965) Fuzzy sets. *Information and Control*, 8, 338–353.