Connotation-differential Prints
Comparing What Is Connoted Through (Fuzzy) Evaluations

Marcelo Loor[1,2] and Guy De Tré[1]
[1]*Dept. of Telecommunications and Information Processing,Ghent University, Sint-Pietersnieuwstraat 41 B - 9000, Ghent, Belgium*
[2]*Dept. of Electrical and Computer Engineering, ESPOL University, Campus Gustavo Galindo V., Km. 30.5 Via Perimetral, Guayaquil, Ecuador*

Keywords: Meaningful Comparison, Semantic Richer Similarity, Qualitative Comparison.

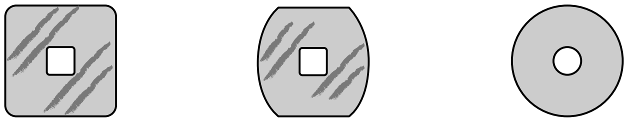
Abstract: To evaluate the level to which an object belongs (or not) to a particular set, say A, one could focus on some object's features according to what one understands by A. With this consideration, using the evaluations of a group of objects given by two persons, we want to determine the level to which their individual understandings of A match. Therefore, hypothesizing that a difference in understandings (or connotations) of A could be marked by a difference in one or more of the evaluations, we propose a *connotation-differential print* (CDP). A CDP is a representation of any difference in connotations of A between two persons in a form that makes itself available to computation. Additionally, we study how to use a CDP to extend a similarity measure for intuitionistic fuzzy sets in order to reach a meaningful comparison between two of them.

1 INTRODUCTION

Imagine the following situation: three cousins, Alice, Bob and Chloe, are evaluating individually to which degree a cookie could be seen or not as a Grandma's cookie. Each cousin has a memory of how looks a Grandma's cookie (see Figure 1), which is used as a referent to evaluate all the cookies depicted in Figure 2. Using a unit interval scale where 1 represents the highest level and 0 the lowest, the cousins have given their evaluations as shown in Table 1. It can be seen from the data in this table that, in some evaluations, adding both a 'yes'-value and a 'no'-value is not necessarily equal to 1. Recording evaluations in this fashion allows the cousins to express any hesitation about their judgments as will be explained in Section 4.2. Within this situation, in this paper we study the following problem: *using the evaluations given by two cousins, how can we determine the level to which the referent cookies used by them match?* This kind of problem is of particular relevance to similarity-related processes where, although two (or more) persons understand the meaning of a request (e.g., is this cookie like a Grandma's cookie?), they may have different understandings of such (e.g., Alice vs. Bob's connotations of a Grandma's cookie depicted in Figure 1) —as will be shown in Section 2, Zadeh (Zadeh, 2013) highlights some ideas about truth and meaning that are related somehow to this problem. Hereafter we assume that it is not feasible (nor practical) for two (or more) persons clarifying their individual understandings of a request. Thus, e.g., if a comparison between evaluations given by two cousins is performed, it is possible that they "match" although the cousins have different understandings (or memories) of a Grandma's cookie. Here, our aim is to detect this kind of pseudo-matching in order to achieve more reliable results in such comparisons.

Straightforwardly, someone may consider that measuring the difference between the evaluation sets given individually by two cousins is a direct measure of the difference between the referent cookies used by them. However, this consideration assumes implicitly that if Alice's referent cookie matches with a particular cookie (which is denoted by an Alice's evaluation) and this particular cookie matches with Chloe's referent cookie (which is denoted by a Chloe's evaluation), then Alice's referent cookie should match with Chloe's, i.e., the similarity between cookies is seen as a transitive relation, and also, as a symmetric relation due to it is also assumed that if Alice's referent matches Chloe's, then Chloe's referent should match Alice's. As will shown in Section 3, following the psychological view of similarity presented by Tver-

Loor M. and De Tré G..
Connotation-differential Prints - Comparing What Is Connoted Through (Fuzzy) Evaluations.
DOI: 10.5220/0005079101270136
In *Proceedings of the International Conference on Fuzzy Computation Theory and Applications* (FCTA-2014), pages 127-136
ISBN: 978-989-758-053-6



(a) *cookie a* (Alice). (b) *cookie b* (Bob). (c) *cookie c* (Chloe).

Figure 1: How looks a Grandma's cookie according to each relative (Grandma's cookie example).

(a) *cookie* 1. (b) *cookie* 2. (c) *cookie* 3. (d) *cookie* 4.

Figure 2: Do these cookies look like a Grandma's cookie? (Grandma's cookie example).

sky (Tversky, 1977) as less stringent but more practical, we consider in this paper that those assumptions should be avoided because they could not reflect the human behavior in comparison judgments. Dealing with such comparison judgments is the theoretical motivation for the paper.

To detect any difference between the referent cookies used by two cousins, we propose using the differences resulting from a comparison of their corresponding evaluation sets in order to obtain a kind of footprint that hints how different (or similar) the referent cookies are; we call this footprint a *connotation-differential print* (CDP). This is based on the following observation. In the abstraction process carried out to give her evaluations, Alice could pay more attention than Chloe to some cookie's features according to what she remembers as a Grandma's cookie, causing a difference in their individual referent objects, which are individual connotations of a Grandma's cookie. We hypothesize that a difference in connotations could be marked by a difference in one or more evaluations. For instance, studying Table 1, it is possible to think that the similarity between Alice and Chloe's connotations (i.e., Alice vs. Chloe) is larger than the similarity between Alice and Bob's connotations (i.e., Alice vs. Bob). Indeed, the evaluations given by Alice and Chloe for *cookie* 1, *cookie* 2 and *cookie* 3 match perfectly; however, the big difference in evaluations given to *cookie* 4 by Alice and Chloe suggests a difference in their individual connotations, in fact, looking into the memories of the cousins depicted in Figure 2, Alice vs. Bob seems more similar than Alice vs. Chloe. This observation is the practical motivation for the paper.

There are two interesting things about a CDP: 1) it is a representation that could be used to determine if any difference between two evaluation sets is caused by a difference in the magnitude of the evaluations, or by a difference in the connotation of the referent objects; and 2) it makes itself available to computation. Thus, e.g., if an intuitionistic fuzzy set (IFS) is used to model the evaluation sets, it is possible to use a CDP to extend a similarity measure for IFS in order to perform a meaningful similarity comparison between two of them (see Section 6.2). As will be shown in Section 7, this contribution could help to overcome an anomaly that happens in some similarity measures for IFSs. Also, the two characteristics could be useful in the following situation. Imagine that Alice needs some help to evaluate additional cookies. Which cousin should be chosen to help her? Maybe the cousin whose connotation of a Grandma's cookie is more similar to Alice's connotation could perform a better evaluation than the cousin with whom Alice agrees exactly on many of the given values. This gives us an idea of a potential application.

Picture yourself as an expert in rating of scripts to be used in language courses for children between 7 and 9 years. Suppose that you have received a huge collection of scripts from several sitcoms to be rated. You have decided to ask the online community for contributions to rate all the scripts in such a huge collection —i.e., you have decided to use a crowdsourcing model. It concerns you that in this kind of model you know nothing about the people who perform the job (which consist in rating one or more scripts), as well as the different understandings that they may have about it. You know that in a crowd-sourced context a contributor's answer could be, among others, affected by personal views, experience or background. Thus, you have written down a request explaining as clearly as possible how each script should be rated, considering that it is not feasible to clarify individually any doubt about the request. Also, you have considered as necessary to use a flexible scale (which is similar to that used in Grandma's cookie example) in order to allow the contributors express any hesitation in their answers. Moreover, you have decided that the contributors must qualify for the job, so they must perform an analysis of some scripts that you have al-

Table 1: To which degree each cookie in Figure 2 is seen or not as a Grandma's cookie by each cousin (Grandma's cookie example).

	yes	*no*
cookie 1	0.6	0.3
cookie 2	0.7	0.3
cookie 3	0.2	0.8
cookie 4	0.9	0.1

(a) Alice.

	yes	*no*
cookie 1	0.4	0.3
cookie 2	0.5	0.3
cookie 3	0	0.9
cookie 4	0.7	0.2

(b) Bob.

	yes	*no*
cookie 1	0.6	0.3
cookie 2	0.7	0.3
cookie 3	0.2	0.8
cookie 4	0.1	0.9

(c) Chloe.

ready analyzed. It is quite important for you being as sure as possible that the contributors perform an analysis similar to yours. If you can find anyone who performs an analysis connoting what you connote in yours, you could trust his or her jobs more than others. As you may have already noticed, our proposal could help you to choose your right contributors and to assess the quality of the jobs that they will perform.

To describe how our proposal provides an answer to the *Grandma's cookie* example, this paper is structured as follows. Section 2 presents some related works. Section 3 shows why the paper considers the psychological view of similarity presented in (Tversky, 1977). Section 4 shows how the intuitionistic fuzzy set (IFS) concept (Atanassov, 1986; Atanassov, 2012) could be used to model each of the evaluation sets given by the cousins. Section 5 uses the spot-difference concept (Loor and De Tré, 2013) to figure out the difference in evaluations given for each cookie by two cousins. Section 6 introduces and explains the novel concept of connotation-differential print (CDP) and shows how it could be used to extend a similarity measure for IFS in order to perform meaningful comparisons. Section 7 compares our meaningful similarity measure with others. We conclude the paper in Section 8.

2 RELATED WORKS AND DISCUSSIONS

About detecting any difference between two referents used by two persons, who are evaluating to which degree an object could be considered to be similar to such a referent, we found in (Zadeh, 2013) some ideas of Zadeh about truth and meaning that are related somehow to ours. First, we agree that it is necessary for two persons, P_1 and P_2 to understand the meaning of a proposition p to assess its truth value —e.g., Alice and Bob should understand what is a Grandma's cookie to evaluate the level to which *cookie i* is seen as such. Second, we agree that understanding the meaning of p is not sufficient for P_1 and P_2 to determine to which level their individual assessments of p match —e.g., although both Alice and Bob understand individually what a Grandma's cookie is, this is not enough to determine if both cousins have the same connotation of such a cookie. Finally, we agree that what is needed is a representation of any difference about the meaning of p between P_1 and P_2 in a form that lends itself to computation —this is why we propose a *connotation-differential print*.

About the semantic interpretation of the elements of an IFS (see Section 4), we found in the theoretical model proposed by Ekman in (Ekman, 1963) — which is used to compare two perceptions from an observer— some analogies that fit with those used in our vector based interpretation of the membership and non-membership of such elements. In his model, Ekman considers that the perceptual intensity is depicted by the magnitude of a vector, and the perceptual quality, by the vector's direction. Analogically, the perceptual intensity corresponds to the degree to which one element, say x_i, belongs or not to an IFS A, i.e., $\mu_A(x_i)$ and $\nu_A(x_i)$ respectively; while the perceptual quality corresponds to the connotation of belonging or not to A.

3 CHOOSING ONE SIMILARITY APPROACH

The aim of this section is to show why the paper considers the psychological view of similarity presented by Tversky in (Tversky, 1977) in order to measure any difference between the referent objects used by two cousins in their evaluations.

3.1 Similarity According to Tversky

In (Tversky, 1977), Tversky provides empirical evidence for asymmetric similarities and argues that similarity should not be treated as a symmetric relation. Representing an object in terms of its qualitative features, he considers that the assessment of similarity may be better described as a *comparison of features* rather than as the computation of a metric distance between points —a metric distance function, d, is defined as a scale that assigns to every pair of points, (x,y), a nonnegative number in accord with the following three axioms: $d(x,y) \geq d(x,x)$ (minimality); $d(x,y) = d(y,x)$ (symmetry); $d(x,y) + (y,z) \geq d(x,z)$ (the triangle inequality). He sustains that *similarity judgments* can be regarded as extensions of *similarity statements*, i.e., statements of the form "*x* is like *y*". Such a statement is *directional* —*x* is the *subject*, and *y* is the *referent*— and, in general, it is not equivalent to the converse similarity statement "*y* is like *x*".

In one of his experiments, called *similarity of countries*, a group of people was asked to choose which of two phrases they preferred to use: "country A is similar to country B," or "country B is similar to country A". Tversky observed that most of the people chose the phrase in which the more prominent country served as the referent —e.g., most of the people chose "North Korea is similar to Red China" instead "Red China is similar to Korea." According to him, those results demonstrate the presence of marked asymmetries in the choice of similarity statements. Using what he observed, Tversky presents the following example that casts some doubts on the psychological validity of transitivity in similarity relations: if "Jamaica is similar to Cuba" and "Cuba is similar to Russia" then the phrase "Jamaica is similar to Russia" should be possible, but Jamaica and Russia are not similar at all. He pointed out that in "Jamaica is similar to Cuba" has been relevant the geographical proximity, while in "Cuba is similar to Russia" has been relevant their political affinity. This human behavior in which each person considers as relevant any particular feature in order to provide an answer, is a key component of his approach.

As seen in Section 1, there are situations where it is not feasible (nor practical) to clarify individually all the constraints in a request. Thus, even though two or more persons understand the meaning of a request, they may have different understandings of it according to what they consider as relevant. Due to this kind of human behavior was considered by Tversky in his approach of similarity, we take it into account in the paper.

4 MODELING (FUZZY) EVALUATIONS

For the purpose of modeling the evaluations given by each cousin, this section introduces the intuitionistic fuzzy set (IFS) concept and uses a semantic interpretation of its components in order to model the components of the *Grandma's cookie* example.

4.1 IFS Concept

As an extension of a fuzzy set (Zadeh, 1965), an IFS A^* in E (Atanassov, 1986; Atanassov, 2012) is defined as a collection such that

$$A^* = \{\langle x_i, \mu_A(x_i), \nu_A(x_i)\rangle | x_i \in E\} \tag{1}$$

where sets E and A are considered to be fixed, $A \subset E$, functions $\mu_A : E \to [0,1]$ and $\nu_A : E \to [0,1]$ define the *degree of membership* and the *degree of non-membership* of $x_i \in E$ to the set A respectively, and for each element $x_i \in E$

$$0 \leq \mu_A(x_i) + \nu_A(x_i) \leq 1. \tag{2}$$

The lack of knowledge about the membership (or non-membership) of element $x_i \in E$ to set A is expressed by

$$\pi_A(x_i) = 1 - \mu_A(x_i) - \nu_A(x_i) \tag{3}$$

and it is defined as the *degree of non-determinacy* —also known as *hesitation margin* (Szmidt and Kacprzyk, 2013).

4.2 A Semantic Interpretation of Components of an IFS

Using a semantic interpretation of IFS's components, we model the components of the *Grandma's cookie* example as follows. The collection of all cookies depicted in Figure 2 corresponds to the set E, where each cookie is denoted by x_i and $i = 1,2,3,4$. A group of cookies considered by Alice to be a Grandma's cookie corresponds to the set A (i.e., $A \subset E$). The degree to which *cookie i* is considered by Alice to be a Grandma's cookie is represented by $\mu_A(x_i)$ — here, $\mu_A(x_i)$ is interpreted as a *degree of similarity* between x_i and r_A, where r_A is what Alice remembers as a Grandma's cookie (cf. the semantics of the membership grades treated in (Dubois et al., 2000)). The degree to which *cookie i* is considered by Alice not to be a Grandma's cookie is represented by $\nu_A(x_i)$. The degree to which Alice hesitates if *cookie i* is considered or not to be a Grandma's cookie is represented by $\pi_A(x_i)$. In this way, the evaluations given by Alice correspond to the IFS $A^* = \{\langle x_i, \mu_A(x_i), \nu_A(x_i)\rangle | x_i \in E\}$. Likewise, the group of cookies considered by Bob and Chloe to be a Grandma's cookie correspond to the sets B and C respectively (i.e., $B \subset E$ and $C \subset E$), and their evaluations, to the IFSs $B^* = \{\langle x_i, \mu_B(x_i), \nu_B(x_i)\rangle | x_i \in E\}$ for Bob and $C^* = \{\langle x_i, \mu_C(x_i), \nu_C(x_i)\rangle | x_i \in E\}$ for Chloe.

It is important to note that, although the definition shows the difference between the IFS A^* and the set

A, for simplicity (Atanassov, 1986; Atanassov, 2012) the expression (1) will be denoted by

$$A = \{\langle x_i, \mu_A(x_i), \nu_A(x_i)\rangle | x_i \in E\}, \quad (4)$$

in the remainder of the paper. Hence, the evaluation sets will be denoted by $A = \{\langle x_i, \mu_A(x_i), \nu_A(x_i)\rangle | x_i \in E\}$ for Alice, $B = \{\langle x_i, \mu_B(x_i), \nu_B(x_i)\rangle | x_i \in E\}$ for Bob, and $C = \{\langle x_i, \mu_C(x_i), \nu_C(x_i)\rangle | x_i \in E\}$ for Chloe.

5 COMPARING EVALUATIONS VS. COMPARING WHAT THEY COULD MEAN

We have already mentioned in the theoretical and practical motivation for the paper (see Section 1) that measuring the similarity of the evaluation sets given individually by two cousins is not a good option to measure the similarity of the referent cookies because of the implicit assumption about transitivity and symmetry in such a comparison. Therefore, although we have modeled the evaluation sets as IFSs, in this paper we should avoid using any of the similarity measures for IFSs that assume the similarity to be a dual notion of a metric distance (cf. the similarity measures presented in Section 7.1).

Considering what Tversky proposed in (Tversky, 1977) about directionality, asymmetry and not transitivity in comparison judgments, this section introduces some concepts presented in (Loor and De Tré, 2013) in order to know not just the magnitude, but also the meaning (or sense) behind a comparison between two IFSs.

5.1 Spot-difference Concept

As seen in Section 4.2, from a semantic point of view, $\mu_A(x_i)$ and $\nu_A(x_i)$ represent the extent (or magnitude) to which *cookie i* is considered by Alice respectively to be a Grandma's cookie, and not to be so. This reflects that $\mu_A(x_i)$ and $\nu_A(x_i)$ have different meanings (or directions), therefore, such an evaluation of *cookie i* given by Alice could be interpreted as a vector

$$\mathbf{a_i} = \begin{pmatrix} \mu_A(x_i) & + & \alpha_A \cdot \pi_A(x_i) \\ \nu_A(x_i) & + & (1-\alpha_A) \cdot \pi_A(x_i) \end{pmatrix}, \quad (5)$$

where $\alpha_A \in [0,1]$ is considered to be a *hesitation splitter* from Alice (Loor and De Tré, 2013) –hereby, the hesitation splitter α_A allows it to split any Alice's hesitation about seen *cookie i* as a Grandma's cookie between $\mu_A(x_i)$ and $\nu_A(x_i)$.

To compare the evaluations for *cookie i* given by Alice and Bob, which are interpreted respectively as vectors $\mathbf{a_i}$ and $\mathbf{b_i}$, the area of the parallelogram formed by both vectors could be used as a reference to measure the difference between them. Thus, the larger this area, the larger the difference between $\mathbf{a_i}$ and $\mathbf{b_i}$. Within this approach, the largest area is given by the vectors $\mathbf{m_f} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{n_f} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Definition 1 (Spot-difference)**.** *Consider an element $x_i \in E$. Let $\mathbf{a_i}$ be a vector representing the membership and non-membership of x_i to the IFS A and $\mathbf{b_i}$ a vector representing the membership and non-membership of x_i to the IFS B. A measure of their differences is known as* spot-difference *and is defined by*

$$dif(\mathbf{a_i}, \mathbf{b_i}) = \frac{\mathbf{a_i} \times \mathbf{b_i}}{\mathbf{m_f} \times \mathbf{n_f}}, \quad (6)$$

where $\times$ denotes the vector product (Loor and De Tré, 2013).

An expression obtained from (6) is

$$dif(\mathbf{a_i}, \mathbf{b_i}) = (\mu_A(x_i) - \mu_B(x_i)) + (\alpha_A \cdot \pi_A(x_i) - \alpha_B \cdot \pi_B(x_i)), \quad (7)$$

which could be semantically interpreted as follows. The first part of the expression, $(\mu_A(x_i) - \mu_B(x_i))$, denotes that the difference between Alice and Bob's evaluations of *cookie i* is determined partly by the difference in levels to which *cookie i* is considered by them to be a Grandma's cookie. The second part, $(\alpha_A \cdot \pi_A(x_i) - \alpha_B \cdot \pi_B(x_i))$, denotes that the difference is also influenced by any doubt about considering *cookie i* to be a Grandma's cookie, and moreover, this part could be affected by managing both Alice (α_A) and Bob (α_B)'s hesitation splitters —a managing strategy could be applying the same rule for Alice and Bob, thus, $\alpha_A = \alpha_B = \alpha$, in which case (7) could be expressed as

$$dif^{\alpha}(\mathbf{a_i}, \mathbf{b_i}) = (\mu_A(x_i) - \mu_B(x_i)) + \alpha(\pi_A(x_i) - \pi_B(x_i)). \quad (8)$$

The $(+/-)$ sign in a spot-difference result semantically denotes the *relative difference* between Alice and Bob's evaluations (cf. Tversky's consideration about directionality for comparison judgments (Tversky, 1977)). For example, $dif^0(\mathbf{a_1}, \mathbf{b_1}) = +0.2$ means that Alice's evaluation considers *cookie* 1 to be a Grandma's cookie 0.2 *more* than Bob's evaluation; on the other hand, $dif^0(\mathbf{b_1}, \mathbf{a_1}) = -0.2$ means that Bob's evaluation considers *cookie* 1 to be a Grandma's cookie 0.2 *less* than Alice's evaluation.

5.2 Spot-differences Footprint and Similarity

A visual representation of the relative notion of difference given by the spot-difference concept allows it to observe the internal composition of the difference between two evaluation sets (i.e., two IFSs) given individually by two cousins. For instance, consider using (8) with $\alpha = 0$. Figure 3(a) shows the difference between Alice's evaluations and Bob's evaluations for each cookie, as seen from Alice's point of view. Figure 3(b) does so, comparing Chloe's evaluations. In a similar way, Figure 3(c) and Figure 3(d) show the differences between Chloe's evaluations and respectively Alice's and Bob's evaluations, as seen from Chloe's point of view. In these figures, each spot-difference is represented by a ruler of height one marked of with "difference"-units. The black region denotes the magnitude of the corresponding spot-difference, and the position of the black region, above or below the line that represents no-difference, denotes its relative difference. This kind of representation is called a *spot-differences footprint* (Loor and De Tré, 2013).

A spot-differences footprint allows it to obtain a measure for the similarity of two IFSs following the Tversky's *ratio model* (Tversky, 1977). Hereby, the similarity is expressed as a proportion between the common and the distinctive features in a normalized form. The measure presented in (Loor and De Tré, 2013) is defined by

$$S^{\alpha}(A,B) = 1 - \frac{1}{n}\sum_{i=1}^{n}|dif^{\alpha}(\mathbf{a_i},\mathbf{b_i})|, \tag{9}$$

and denotes the degree of similarity $S^{\alpha}(A,B)$ between IFSs A and B. As it could be noted, this expression by itself just represents the magnitude of the similarity between two IFSs and, by using it, it is not possible to conclude about the correspondence in appearance of the two IFSs. For instance, using this expression, the similarity between Alice and Bob is $S^0(A,B) = 0.8$, as well as, the similarity between Alice and Chloe is $S^0(A,C) = 0.8$, i.e., although they differ in their spot-differences footprint (see Figure 3), both Alice vs. Bob and Alice vs. Chloe have the same value as a measure of similarity.

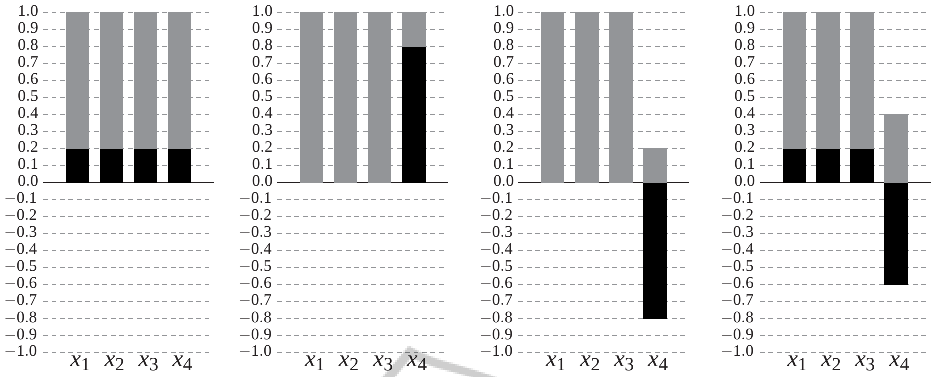
To distinguish between the *"0.8-uniform"* similarity in Alice vs. Bob and the *"0.8-with-a-peak"* similarity in Alice vs. Chloe, it has been proposed in (Loor and De Tré, 2013) to make use of the corresponding spot-differences footprint to extend (9) in order to capture the semantic meaning (or sense) within comparisons. This is the motivation for a *semantic footprint* that is proposed in the next Section.

6 CONNOTATION-DIFFERENTIAL PRINT

At this point, the question is *how could we use a spot-differences footprint to compare what is connoted through evaluations given by two cousins?* In (Tversky, 1977), using a set-theoretical approach, Tversky described the similarity between two objects o_1 and o_2, which are represented as a collection of features O_1 and O_2 respectively, as a *feature-matching* process. Considering this approach, each cookie in *Grandma's cookie* example could be treated as an object with features such as a square shape, linear icing, or with a square hole. Thus, in the abstraction process carried out to give her evaluation, Alice could pay more attention than Bob or Chloe to some cookie's features according to what she remembers as a Grandma's cookie, causing a difference in their individual connotations (of a Grandma's cookie). For instance, let us take a look into each cousin's memory of a Grandma's cookie depicted in Figure 1: Alice's memory is a square cookie with linear icing and a square hole; Bob's memory is a square-round cookie with linear icing and a square hole; and Chloe's memory is a round cookie with no icing and a round hole. When Alice evaluated to which degree *cookie* 1 (see Figure 2(a)) could be seen as a Grandma's cookie, she judged it as 0.6 for 'yes' and 0.3 for 'no' (see Table 1) —it seems that she paid more attention to the square shape of the cookie (which is a common feature between *cookie* 1 and the cookie in her memory) and not to the missing linear icing nor the missing square hole (cf. Tversky's example in (Tversky, 1977)). When Chloe did so, she also judged it as 0.6 for 'yes' and 0.3 for 'no' (see Table 1) — it seems that she paid more attention to the missing icing and not to the square shape nor the missing round hole. Looking at the spot-differences footprint resulting from Alice-vs.-Chloe comparison (see Figure 3(b)), Alice could hint that, although they agree on evaluations for *cookie* 1, *cookie* 2 and *cookie* 3, Chloe paid less attention to the square shape or square hole of *cookie* 4, i.e., Alice could hint that her connotation of a Grandma's cookie differs from Chloe's. This illustrates how spot-differences footprints can help to compare what is connoted through evaluations.

6.1 Connotation-differential Marker

Following the aforementioned idea, a way to detect the difference between Alice's connotation of a Grandma's cookie in comparison to Bob's or Chloe's is by shifting the focus onto some cookies having fea-



(a) Alice vs. Bob. (b) Alice vs. Chloe. (c) Chloe vs. Alice. (d) Chloe vs. Bob.

Figure 3: Comparing spot-differences footprints.

tures that she considered to be (or not) representative in a Grandma's cookie. For example, Alice could focus on *cookie* 4 (see Figure 2(d)) because it has the same square shape, as well as the same square hole, as her memory of a Grandma's cookie (see Figure 1(a)), and *cookie* 3 (see Figure 2(c)) because it has neither the square shape nor the square hole remembered by her —following these assumptions, *cookie* 4 and *cookie* 3 have respectively the best and the worst of Alice's evaluations (see Table 1). Focusing to one of these "representative" cookies, we propose to use a marker that denotes how a spot-difference related to this cookie should be managed, i.e., if a spot-difference should be treated as matching on evaluations or not. Recalling the semantic interpretation in a spot-difference result, such a marker should reflect the relative difference between two evaluations, therefore, we define it as follows.

Definition 2 (Connotation-Differential Marker)**.** *Consider an element $x_i \in E$. Let $\mathbf{a_i}$ be a vector representing the membership and non-membership of x_i to the IFS A, $\mathbf{b_i}$ a vector representing the membership and non-membership of x_i to the IFS B, and $dif^{\alpha}(\mathbf{a_i},\mathbf{b_i})$ a spot-difference between $\mathbf{a_i}$ and $\mathbf{b_i}$. Now consider a set $S = \{$ϕ, ⫯, ⫰$\}$. A connotation-differential marker (CDM) is a symbol $s \in S$ that denotes the relative amount of difference between $\mathbf{a_i}$ and $\mathbf{b_i}$ according to the following conditions:*

- *if $|dif^{\alpha}(\mathbf{a_i},\mathbf{b_i})| \leq \delta$ then $s =$ ϕ,*
- *if $dif^{\alpha}(\mathbf{a_i},\mathbf{b_i}) > \delta$ then $s =$ ⫯,*
- *if $dif^{\alpha}(\mathbf{a_i},\mathbf{b_i}) < \delta$ then $s =$ ⫰,*

where $\delta \in [0,1]$.

In this way, for instance, from Alice's point of view, with $\delta = 0.2$, the spot-difference for *cookie* 4 between her and Bob, $dif^0(a_4,b_4) = 0.2$, could be treated as a matching on evaluations and denoted by ϕ. By contrast, the spot-difference for *cookie* 4 between Alice and Chloe, $dif^0(a_4,c_4) = 0.8$, could be treated as a difference in evaluations where Alice considers that *cookie* 4 is more similar to a Grandma's cookie and, as seen from her point of view, it is denoted by ⫯.

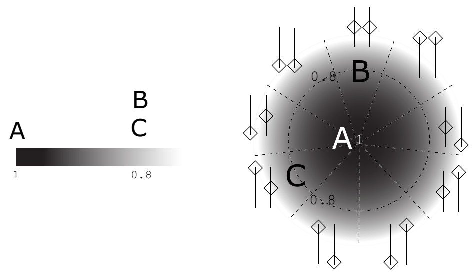
6.2 A Meaningful Comparison

Knowing what a CDM could denote individually, we could put one or more CDMs together in order to obtain a representation that hints if a cousin has paid attention or not to the same cookie's features that have been focused by another during the evaluation process. A way to do that is by placing one or more CDMs in a sequence with a particular order. For example, to perform the comparisons from Alice's point of view, i.e., considering as referent what Alice remembers as a Grandma's cookie, we could build sequences with two CDMs: the first one related to her best evaluated cookie, i.e., *cookie* 4 (see Table 1), and the second one, to her worst evaluated cookie, i.e., *cookie* 3; thus, with $\delta = 0.2$, a sequence corresponding to (the comparison) Alice vs. Bob (see Figure 3(a)) is "ϕϕ", and one corresponding to Alice vs. Chloe (see Figure 3(b)) is "⫯ϕ". Despite using only two CDMs, looking at Alice-vs.-Bob sequence, Alice can now distinguish that Bob seems to agree with her about *cookie* 4 having one or more features to consider it to be a Grandma's cookie, as well as, *cookie* 3 having one or more features to consider it not to be so. On the other hand, looking at the first CDM in Alice-vs.-Chloe sequence, Alice could become aware that some features of *cookie* 4 make Chloe to consider it not to be a Grandma's cookie and, thus, Alice could hint that she and Chloe have used different connotations (of a Grandma's cookie) as referents. Now, let us perform the comparison from Chloe's point of view, i.e., considering as referent what Chloe remembers as Grandma's cookie. Using the same strategy to build a sequence, the first CDM is related to the Chloe's best evaluated cookie, i.e. *cookie* 2 (see Table 1), and the second to her worst's, i.e., *cookie* 4. Thus, a sequence corresponding to Chloe-vs.-Bob comparison (see Figure 3(c)) is “”, and a sequence corresponding to Chloe-vs.-Alice comparison (see Figure 3(d)) is “”. Looking at these sequences, Chloe could hint that neither Bob nor Alice remember a Grandma's cookie with no icing as she does. Due to these sequences allow Alice or Chloe to hint about a difference in connotations of a Grandma's cookie, we call any of them a *connotation-differential print* (CDP).

As it could be noticed above, a CDP depends on the individual point of view of each cousin. In fact, Alice has chosen the CDMs corresponding to *cookie* 4 and *cookie* 3, while Chloe has chosen *cookie* 2 and *cookie* 4. This is an example of directionality and asymmetry in comparison judgments pointed out by Tversky in (Tversky, 1977). Furthermore, Alice could assign a weight to each CDP in order to determine which cousin's referent is more similar to hers. For example, according to her strategy to build a CDP, denotes a good match, thus, she assigns 1.0 to it. and denote a not too bad match (they could become by increasing δ) therefore, she gives 0.75 to them. , , and denote a big difference, so, she gives 0.25 to them. Finally, and denote a huge difference, so, she assigns 0 to them.

At this point, we could use a CDP to extend (9) to perform a meaningful similarity comparison between two IFSs and, therefore, to achieve better fine-tuned and more reliable results. As an analogy, if (9) tells us about how far is A from B, the use of a CDP tells us in which “direction” B is in relation to A. With this analogy, Figure 4 shows the similarity between evaluations given by Alice and Bob, $S^0(A,B) = 0.8$, and so between Alice and Chloe's, $S^0(A,C) = 0.8$. In Figure 4(a), using just the magnitude, there is no difference between both similarities; in contrast, in Figure 4(b), using the magnitude plus a CDP, it is noteworthy how the “direction” of Alice vs. Bob differs from Alice vs. Chloe's. Hence, from Alice's point of view, the extended similarity between her evaluations and Bob's is $\langle 0.8, \rangle$, and the corresponding one in comparison to Chloe's evaluations is $\langle 0.8, \rangle$. Moreover, using the weights that Alice proposed earlier, we could say that Alice-vs.-Bob's $\langle 0.8, \rangle$ is $0.8 \cdot 1 = 0.8$, and Alice-vs.-Chloe's $\langle 0.8, \rangle$ is $0.8 \cdot 0.25 = 0.2$. This reflects that, considering the individual connotation of a Grandma's cookie, the similarity between Alice and Bob's evaluations is better.

Using the previous analogy, it is also possible to illustrate how using a CDP as supplement could denote in a better way an observed similarity. In the example



(a) (b)

Figure 4: Alice vs. Bob and Alice vs. Chloe similarities.

presented by Tversky in (Tversky, 1977) it is stated that: “considering the similarity between countries, Jamaica is similar to Cuba (because of geographical proximity), Cuba is similar to Russia (because of their political affinity), but Jamaica and Russia are not similar at all”, one may notice that both comparisons, Jamaica vs. Cuba and Cuba vs. Russia, have similar magnitudes in their corresponding similarity measures, but different connotations: the first one focusing on a “geographical proximity”-feature, and the second one, on a “political affinity”-feature. Using such a representation of the difference in connotations, one may observe the reason why Jamaica and Russia are not similar at all. Maybe, someone might argue here that the same properties should be used in both comparisons. However, as was already mentioned in Section 1, it is not feasible (nor practical) in some contexts (e.g., when a crowdsourcing model is used) clarifying all the properties that an object may have. Indeed, this is why (fuzzy) human evaluations are needed in those contexts.

7 AN ADVANTAGE

This section compares our meaningful similarity measure with others in order to show how it could overcome some difficulties such as those pointed out in (Szmidt and Kacprzyk, 2013).

7.1 Some Difficulties in Geometric Similarity Measures for IFSs

Using a geometric interpretation of the three terms in an IFS-element (i.e., the membership, non-membership and hesitation margin), the similarity between two IFSs is usually assumed to be a dual notion of a metric distance (Szmidt and Kacprzyk, 2013). Thus, given two IFSs $A, B \in X$ and a normalized metric distance function $l : X^2 \mapsto [0,1]$, the similarity S between A and B is expressed as $S(A,B) = 1 - l(A,B)$, where l follows the axioms of minimality, symmetry and the triangle inequality (see Section 3.1). In (Szmidt and Kacprzyk, 2013), Szmidt and Kacprzyk have examined the effects of this assumption. They found some difficulties that, according to them, are a result of, first, the symmetry of the three terms, and second, of the important role played by these terms in the definition of the complement of IFSs, which should be considered in such similarity measures. Among others, they have studied the following similarity measures:

$$\begin{aligned} S_H(A,B) = 1 - \frac{1}{2n}\sum_{i=1}^{n} (&|\mu_A(x_i) - \mu_B(x_i)| \\ &+ |\nu_A(x_i) - \nu_B(x_i)| \\ &+ |\pi_A(x_i) - \pi_B(x_i)|) \end{aligned} \tag{10}$$

and

$$\begin{aligned} S_{H2D}(A,B) = 1 - \frac{1}{2n}\sum_{i=1}^{n} (&|\mu_A(x_i) - \mu_B(x_i)| \\ &+ |\nu_A(x_i) - \nu_B(x_i)|) \end{aligned} \tag{11}$$

based on Hamming distance (Szmidt and Kacprzyk, 2000);

$$\begin{aligned} S_e(A,B) = 1 - \Bigg(\frac{1}{2n}\sum_{i=1}^{n} (&(\mu_A(x_i) - \mu_B(x_i))^2 \\ &+ (\nu_A(x_i) - \nu_B(x_i))^2 \\ &+ (\pi_A(x_i) - \pi_B(x_i))^2)\Bigg)^{\frac{1}{2}} \end{aligned} \tag{12}$$

and

$$\begin{aligned} S_{e2D}(A,B) = 1 - \Bigg(\frac{1}{2n}\sum_{i=1}^{n} (&(\mu_A(x_i) - \mu_B(x_i))^2 \\ &+ (\nu_A(x_i) - \nu_B(x_i))^2)\Bigg)^{\frac{1}{2}} \end{aligned} \tag{13}$$

based on Euclidean distance (Szmidt and Kacprzyk, 2000); and, the cosine similarity measure

$$\begin{aligned} S_{mult}(A,B) = \frac{1}{n}\sum_{i=1}^{n} (&\mu_A(x_i)\,\mu_B(x_i) + \nu_A(x_i)\,\nu_B(x_i) \\ &+ \pi_A(x_i)\,\pi_B(x_i)) \,/\, \Big(\mu_A(x_i)^2 + \nu_A(x_i)^2 \\ &+ \pi_A(x_i)^2\Big)^{\frac{1}{2}} \Big(\mu_B(x_i)^2 + \nu_B(x_i)^2 \\ &+ \pi_B(x_i)^2\Big)^{\frac{1}{2}}. \end{aligned} \tag{14}$$

One of the difficulties is exemplified as follows. Consider $X = \{x_0\}$ and the IFSs $M = \{\langle x_0, 1, 0\rangle\}$, $N = \{\langle x_0, 0, 1\rangle\}$ and $H = \{\langle x_0, 0, 0\rangle\}$. Also consider the similarity measure S. If S is (10), (12) or (14), it is obtained that $S(M,N) = 0$ and $S(M,H) = 0$ though N and H are different. This anomaly is generalized to IFSs such as $K = \{\langle x_0, 0.5, 0.3\rangle\}$ and $L = \{\langle x_0, 0.5, 0.2\rangle\}$ where the exchange of "the places" between the non-membership value and the hesitation margin in K and L results in $S(M,K) = S(M,L)$. Due to this anomaly is caused by the symmetry between the non-membership value and the hesitation margin, Szmidt and Kacprzyk also verified the results using the "two terms"-distances (11) and (13). However, they found that the situation does not change in the sense of the information obtained. As a solution to this kind of anomalies, in (Loor and De Tré, 2013), it was proposed (9). Thus, in this example, $S^\alpha(M,N) = 0$ and $S^\alpha(M,H) = \alpha$, which makes sense according to the semantic interpretation given in Section 5.1. Despite this, there is an anomaly that (9) could not manage by itself and it is needed a CDP.

7.2 An Anomaly That a CDP Could Help to Solve

Consider $X = \{x_0\}$ and the IFSs $P = \{\langle x_0, 0.5, 0.5\rangle\}$, $Q = \{\langle x_0, 0.9, 0.1\rangle\}$ and $R = \{\langle x_0, 0.1, 0.9\rangle\}$. Also consider the similarity measure S. If S is (10), (11), (12), (13), (14) or even (9), it is obtained that $S(P,Q) = S(P,R)$ though Q and R are obviously different. This anomaly could be generalized to IFSs such as $V = \{\langle x_0, 0.7, 0.3\rangle\}$ and $W = \{\langle x_0, 0.3, 0.7\rangle\}$ where giving the values $\mu_V(x_0) = \nu_W(x_0)$, $\mu_W(x_0) = \nu_V(x_0)$ and $\pi_V(x_0) = \pi_W(x_0) = 0$ results in $S(P,V) = S(P,W)$. To solve this, we use the extended version of (9) as follows. From (5) the corresponding vector interpretations for x_0 are $\mathbf{p_0} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$, $\mathbf{q_0} = \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix}$ and $\mathbf{r_0} = \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix}$ —notice that $\pi_P(x_0) = 0$, $\pi_Q(x_0) = 0$ and $\pi_R(x_0) = 0$, which means that a hesitation splitter is not necessary. Then, considering the point of view of P and using (8), the spot-differences for x_0 are $dif^\alpha(\mathbf{p_0}, \mathbf{q_0}) = -0.4$ and $dif^\alpha(\mathbf{p_0}, \mathbf{r_0}) = 0.4$. With $\delta = 0.2$, the corresponding connotation-differential marker for $dif^\alpha(\mathbf{p_0}, \mathbf{q_0})$ is $\underset{\diamond}{|}$ and the corresponding one for $dif^\alpha(\mathbf{p_0}, \mathbf{r_0})$ is $\overset{\diamond}{|}$. To build the CDPs for P-vs.-Q and P-vs.-R comparisons, we use the connotation-differential markers given for x_0, thus, from P's view, $\underset{\diamond}{|}$ is a CDP for P-vs.-Q, and $\overset{\diamond}{|}$ is a CDP for P-vs.-R. Finally, from (9) we obtain $S^\alpha(P,Q) = 0.6$ and $S^\alpha(P,R) = 0.6$, and, using the corresponding CDPs, we extend them to $\langle 0.6, \underset{\diamond}{|}\rangle$ and $\langle 0.6, \overset{\diamond}{|}\rangle$ respectively. As expected, $\langle 0.6, \underset{\diamond}{|}\rangle$ and $\langle 0.6, \overset{\diamond}{|}\rangle$ are different. One might argue that if the same weight is assigned to

$\underset{\diamond}{|}$ and $\overset{\diamond}{|}$, then $S(P,Q) = S(P,R)$; however, it will depend on the semantic interpretation of who assigns the weights, which could be important in a decision making context.

8 CONCLUSIONS

Following the psychological view of similarity presented by Tversky in (Tversky, 1977), we have envisaged an object as a collection of features and considered that, when a person is evaluating to which level an object belongs (or not) to a particular set, say A, she or he could focus on some specific features according to what she or he understands by A, i.e., by focusing on such features, this person makes an personal representation of an object that is used as referent of A. Thus, even if two persons understand what is A, they could have different connotations of it. Then, we have hypothesized that, if two persons have evaluated to which degree each object within a group could be considered to be similar to A, a difference in connotations of A could be marked by a difference in one or more of their evaluations. Therefore, modeling such evaluations using a semantic interpretation of the concept of intuitionistic fuzzy set given by Attanasov in (Atanassov, 1999; Atanassov, 2012), we have presented a *connotation-differential marker* to denote how a difference between two (fuzzy) evaluations for an object given by two persons could be managed.

Choosing one of the points of view of the persons who perform the evaluation, we have shown how to build a sequence from connotation-differential markers corresponding to one or more representative objects for this evaluator. We call this sequence a *connotation-differential print*, which is assumed to be a representation of the difference in connotations of A between the person whose point of view is chosen and any other person. Furthermore, we have given an example of how a person could assign a weight to each of the possible connotation-differential prints according to the strategy followed by this person to build such sequences.

We have illustrated how to use a connotation-differential print as a supplement of a similarity measure for intuitionistic fuzzy sets in order to perform a meaningful comparison between two of them. Finally, we have compared our extended similarity measure with others in order to show one of its possible advantages.

In our future work, we will further explore the applicability of connotation-differential prints in order to perform qualitative comparisons. In particular, it interests us to look into the following problems: 1) assessing the quality of data gathered in a crowd-sourced context; and, 2) detecting any difference in understandings of a given topic between two knowledge databases that come from different sources and need to be merged.

REFERENCES

Atanassov, K. (1999). Interval valued intuitionistic fuzzy sets. In *Intuitionistic Fuzzy Sets*, volume 35 of *Studies in Fuzziness and Soft Computing*, pages 139–177. Physica-Verlag HD.

Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy sets and Systems*, 20(1):87–96.

Atanassov, K. T. (2012). *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Dubois, D., Ostasiewicz, W., and Prade, H. (2000). Fuzzy sets: history and basic notions. In *Fundamentals of Fuzzy Sets*, pages 21–124. Springer.

Ekman, G. (1963). A direct method for multidimensional ratio scaling. *Psychometrika*, 28(1):33–41.

Loor, M. and De Tré, G. (2013). Vector based similarity measure for intuitionistic fuzzy sets. In *New Trends in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics*.

Szmidt, E. and Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 114(3):505–518.

Szmidt, E. and Kacprzyk, J. (2013). Geometric similarity measures for the intuitionistic fuzzy sets. In *8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13)*. Atlantis Press.

Tversky, A. (1977). Features of similarity. *Psychological review*, 84(4):327.

Zadeh, L. (1965). Fuzzy sets. *Information and control*, 8(3):338–353.

Zadeh, L. A. (2013). Toward a restriction-centered theory of truth and meaning (rct). *Information Sciences*, 248(0):1 – 14.