Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 178 (2008) 381–402
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets

Dongrui Wu, Jerry M. Mendel *

Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564, USA

Received 7 October 2006; received in revised form 3 March 2007; accepted 30 April 2007

Abstract

Fuzzy logic is frequently used in *computing with words* (CWW). When input words to a CWW engine are modeled by interval type-2 fuzzy sets (IT2 FSs), the CWW engine's output can also be an IT2 FS, $\widetilde{A}$, which needs to be mapped to a linguistic label so that it can be understood. Because each linguistic label is represented by an IT2 FS $\widetilde{B}_i$, there is a need to compare the similarity of $\widetilde{A}$ and $\widetilde{B}_i$ to find the $\widetilde{B}_i$ most similar to $\widetilde{A}$. In this paper, a vector similarity measure (VSM) is proposed for IT2 FSs, whose two elements measure the similarity in shape and proximity, respectively. A comparative study shows that the VSM gives more reasonable results than all other existing similarity measures for IT2 FSs for the linguistic approximation problem. Additionally, the VSM can also be used for type-1 FSs, which are special cases of IT2 FSs when all uncertainty disappears.

Keywords: Similarity measure; Compatibility measure; Type-1 fuzzy set; Interval type-2 fuzzy set; Computing with words; Linguistic approximation

1. Introduction

Zadeh coined the phrase "*computing with words*" (CWW) [48,49]. According to him, CWW is "*a methodology in which the objects of computation are words and propositions drawn from a natural language*". It is "*inspired by the remarkable human capability to perform a wide variety of physical and mental tasks without any measurements and any computations.*" Nikravesh [35] further pointed out that CWW is "*fundamentally different from the traditional expert systems which are simply tools to 'realize' an intelligent system, but are not able to process natural language which is imprecise, uncertain and partially true.*"

Our thesis is that *words mean different things to different people* and so there is uncertainty associated with words, which means that fuzzy logic must somehow use this uncertainty when it computes with words [25,26].

* Corresponding author. Tel.: +1 213 740 4445; fax: +1 213 740 4651.
E-mail addresses: dongruiw@usc.edu (D. Wu), mendel@sipi.usc.edu (J.M. Mendel).

0020-0255/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2007.04.014

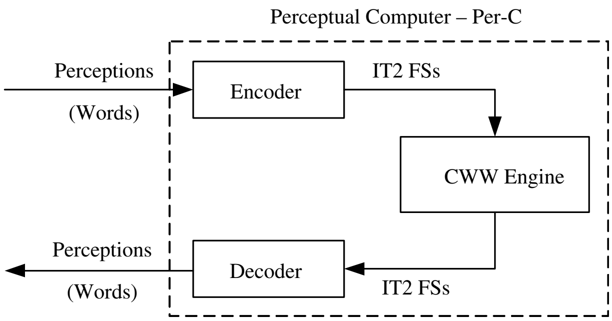

Fig. 1. Conceptual structure of CWW.

Hence, we argue that interval type-2 fuzzy sets (IT2 FSs) should be used in CWW [28]. We will limit our discussions to IT2 FSs in this paper.

A specific architecture is proposed in [27] for making judgements by CWW. A slightly modified architecture is shown in Fig. 1. It will be called a *perceptual computer* – Per-C for short. Perceptions (i.e., granulated terms, words) activate the Per-C and are also output by the Per-C; so, it is possible for a human to interact with the Per-C just using a vocabulary of words. In Fig. 1, the *encoder*[1] transforms linguistic perceptions into IT2 FSs that activate a *CWW engine*. The *decoder*[2] maps the output of the CWW engine into a word. Usually a vocabulary (codebook) is available, in which every word is modeled as an IT2 FS. The output of the CWW engine is mapped into a word (in that vocabulary) most similar to it.

The CWW engine, e.g. rules, the linguistic weighted average (LWA) [43], etc., maps IT2 FSs into IT2 FSs. If the CWW engine is rule-based, its output may be a crisp number (e.g., after defuzzification), in which case the decoder can map this number into a word in the vocabulary, as explained in [27]. On the other hand, if the CWW engine uses the LWA, its output is an IT2 FS $\widetilde{A}$, or if the CWW engine is rule-based, but its output is also an IT2 FS $\widetilde{A}$, then the decoder must also map $\widetilde{A}$ into a word in the vocabulary. In this paper it is assumed that the output of the CWW engine is an IT2 FS $\widetilde{A}$.

How to transform linguistic perceptions into IT2 FSs, i.e. the encoding problem, has been considered in [30–32]. This paper considers the decoding problem, or, as called by Zadeh [48,49], *linguistic approximation*, i.e. how to map an IT2 FS $\widetilde{A}$ into a word (linguistic label). More specifically, given a vocabulary consisting of N words with their associated IT2 FSs $\widetilde{B}_i$ $(i = 1, \dots, N)$, our goal is to find the $\widetilde{B}_i$ which most closely resembles $\widetilde{A}$, the output of the CWW engine. The word associated with that $\widetilde{B}_i$ will then be viewed as the output of the Per-C. To do this, it must be possible to compare the similarity between two IT2 FSs. A vector similarity measure (VSM) for IT2 FSs is proposed in this paper.

The rest of this paper is organized as follows: Section 2 gives the definitions of similarity, proximity and compatibility, which are closely related to each other. Section 3 reviews four existing similarity measures for IT2 FSs. Section 4 proposes a VSM for IT2 FSs. Section 5 provides discussions on a number of issues and shows that the VSM for IT2 FSs can also be used for type-1 (T1) FSs when all uncertainty disappears. Section 6 draws conclusions. Some background material about IT2 FSs is given in Appendix A. Proofs of the theorems are given in Appendix B.

2. Definitions

Similarity, proximity and compatibility are three closely related concepts. There are different definitions on the meanings of them [8,12,20,24,38,45,46]. According to Yager [45], a proximity relationship between two T1 FSs A and B on a domain X is a mapping p: $X \times X \to T$ having the properties: (1) *Reflexivity:* $p(A, A) = 1$; and, (2) *Symmetry:* $p(A, B) = p(B, A)$. Often T is the unit interval.

[1] Zadeh calls this *constraint explicitation* in [48,49]. In [50] and some of his recent talks, he calls this *precisiation*.
[2] Zadeh calls this *linguistic approximation* in [48,49].

A similarity relationship between two FSs A and B on a domain X is a mapping s: $X \times X \to T$ having the properties [45]: (1) *Reflexivity:* $s(A,A)=1$; (2) *Symmetry:* $s(A,B)=s(B,A)$; (3) *Transitivity:* $s(A,B) \geqslant s(A,C) \wedge s(C,B)$, where C is an arbitrary FS on domain X. Observe that here a similarity relationship adds the additional requirement of transitivity, though whether this should be done is still under debate [22].

There are also some weakened forms of transitivity used in the literature, e.g. [6]

Weakened Transitivity Form 1: If $A \leqslant B \leqslant C$,[3] then $s(A,B) \geqslant s(A,C)$ and $s(B,C) \geqslant s(A,C)$.

Let $c(A)$ denote the centroid of a FS A. In this paper the following two even more weakened forms of transitivity are also considered:

Weakened Transitivity Form 2: If A, B and C are of the same shape, and $c(A) \leqslant c(B) \leqslant c(C)$, then $s(A,B) \geqslant s(A,C)$ and $s(B,C) \geqslant s(A,C)$.

Weakened Transitivity Form 3: If $c(A)=c(B)=c(C)$ and $A \leqslant B \leqslant C$, then $s(A,B) \geqslant s(A,C)$ and $s(B,C) \geqslant s(A,C)$.

Compatibility is a wider concept. According to Cross and Sudkamp [8], "*the term compatibility is used to encompass various types of comparisons frequently made between objects or concepts. These relationships include similarity, inclusion, proximity, and the degree of matching.*"

In summary, similarity is included in proximity, and both similarity and proximity are included in compatibility. In this paper we focus on similarity measures; however, some proximity and compatibility measures are also included for comparison purpose.

3. Existing similarity/compatibility measures for IT2 FSs

The literature on similarity/compatibility measures for T1 FSs is quite extensive. According to Bustince [7], "*there are approximately 50 expressions for determining how similar two fuzzy sets are.*" Some commonly used similarity/compatibility measures for T1 FSs are summarized in Table 1 because they will be used by some IT2 FS similarity measures. For more details on T1 similarity/compatibility measures, see [8] and its references. Because compatibility measures for T1 FSs are not the focus of this paper, and there are too many of them, we do not distinguish between compatibility, proximity and similarity in Table 1.

To the best knowledge of the authors, only four similarity/compatibility measures for IT2 FSs have appeared to date, and they are briefly reviewed next. As pointed out by Cross and Sudkamp [8], "*ideally, the selection of a compatibility measure should be justifiable based upon the problem domain, the information being processed, and the inherent properties of the particular measure.*" The four similarity/compatibility measures were originally proposed for different problem domains; however, since the focus of this paper is the linguistic approximation problem in CWW, their ability as a decoder is analyzed.

3.1. Mitchell's IT2 FS similarity measure

Mitchell was the first to define a similarity measure for *general* T2 FSs [34]. For the purpose of this paper, only its special case is explained, when both $\widetilde{A}$ and $\widetilde{B}$ are IT2 FSs:

(1) Discretize the primary variable's universe of discourse, X, into L points, that are used by both $\widetilde{A}$ and $\widetilde{B}$.
(2) Find M embedded T1 MFs (see (A.4) in Appendix A) for IT2 FS $\widetilde{A}$ ($m=1,2,\ldots,M$), i.e.

$$\mu_{A_e^m}(x_l)=r_m(x_l)\times[\overline{\mu}_{\widetilde{A}}(x_l)-\underline{\mu}_{\widetilde{A}}(x_l)]+\underline{\mu}_{\widetilde{A}}(x_l)\quad l=1,2,\ldots,L \tag{1}$$

where $r_m(x_l)$ is a random number chosen uniformly in $[0,1]$, and $\underline{\mu}_{\widetilde{A}}(x_l)$ and $\overline{\mu}_{\widetilde{A}}(x_l)$ are the lower and upper memberships of $\widetilde{A}$ at x_l.
(3) Similarly, find N embedded T1 MFs, $\mu_{B_e^n}$ ($n=1,2,\ldots,N$), for IT2 FS $\widetilde{B}$, i.e.,

$$\mu_{B_e^n}(x_l)=r_n(x_l)\times[\overline{\mu}_{\widetilde{B}}(x_l)-\underline{\mu}_{\widetilde{B}}(x_l)]+\underline{\mu}_{\widetilde{B}}(x_l)\quad l=1,2,\ldots,L \tag{2}$$

[3] $A \leqslant B$ if and only if for $\forall x \in X$, $\mu_A(x) \leqslant \mu_B(x)$.

Table 1
Summary of similarity/compatibility measures for T1 FSs

Similarity/compatibility measure	Equation
Set-theoretic	
Tversky's method [41]	$s_T(A,B)=f(A\cap B)/[f(A\cap B)+a\cdot f(A-B)+b\cdot f(B-A)]$, where f is a function satisfying $f(A\cup B)=f(A)+f(B)$ for disjoint A and B
Jaccard's method [17]	$s_J(A,B)=f(A\cap B)/f(A\cup B)$, where f is defined above
Dubois and Prade's method [10]	$s_D(A,B)=g(\overline{(A\cup B)\cap(\bar{A}\cup\bar{B})})$ or $s_D(A,B)=g(\overline{(\bar{A}\cup B)\cap(A\cup\bar{B})})$, where g satisfies: (1) $g(\emptyset)=0$, (2) $g(X)=1$, and (3) $g(A)\leqslant g(B)$ if $A\subseteq B$
Proximity-based	
• *Minkowski's r-metric based* [52]	$d_r(A,B)\equiv(\sum_{i=1}^n\lvert\mu_A(x_i)-\mu_B(x_i)\rvert^r)^{1/r},\quad r\geqslant 1$ $d_r(A,B)$ is a distance measure
Normalization approach [8]	$s_N(A,B)=1-d_r(A,B)/n$
Conversion function approach [36]	$s_C(A,B)=[1+(d_r(A,B)/s)^t]^{-1}$, where s and t are positive constants
• *Angular coefficient based*	
Bhattacharya's distance [1]	$s_B(A,B)=\dfrac{\sum_{i=1}^n\mu_A(x_i)\cdot\mu_B(x_i)}{\left(\sum_{i=1}^n\mu_A(x_i)^2\right)^{1/2}\cdot\left(\sum_{i=1}^n\mu_B(x_i)^2\right)^{1/2}}$
• *Interval-based*	$A_\alpha\equiv[a_1(\alpha),a_2(\alpha)]$ and $B_\alpha\equiv[b_1(\alpha),b_2(\alpha)]$ are α-cuts on A and B $q(A_\alpha,B_\alpha)\equiv\max(\lvert a_1(\alpha)-b_1(\alpha)\rvert,\lvert a_2(\alpha)-b_2(\alpha)\rvert)$,
Hausdorff distance based [37,52]	$s_H(A,B)=[1+(q_*(A,B)/s)^t]^{-1}$, where $q_*(A,B)$ can be $q(A_1,B_1)$, $\int_0^1 q(A_\alpha,B_\alpha)d\alpha$, or $\sup_{\alpha\geqslant 0} q(A_\alpha,B_\alpha)$
Dissemblance index based [8]	$d(A_\alpha,B_\alpha)\equiv[\lvert a_1(\alpha)-b_1(\alpha)\rvert+\lvert a_2(\alpha)-b_2(\alpha)\rvert]/(2\lvert X\rvert)$, where $\lvert X\rvert$ is the length of the domain of $A\cup B$. $s_D(A,B)=d(A_1,B_1)$, or $\int_0^1 d(A_\alpha,B_\alpha)d\alpha$, or $\sup_{\alpha\geqslant 0} d(A_\alpha,B_\alpha)$
• *Linguistic approximation based*	
Bonissone's method [4]	$s_B(A,B)=[1-\int_X(\frac{\mu_A(x)\mu_B(x)}{\mathrm{card}(A)\cdot\mathrm{card}(B)})^{1/2}\mathrm{d}x]^{1/2}$, where card($A$) [card($B$)] is the cardinality of A (B)
Logic-based	
Hirota and Pedrycz's method [15,16]	$[\mu_A(x_i)\Longleftrightarrow\mu_B(x_i)]\equiv[\mu_A(x_i)\rightarrow\mu_B(x_i)]\wedge[\mu_B(x_i)\rightarrow\mu_A(x_i)]$ $[\mu_A(x_i)=\mu_B(x_i)]\equiv\{[\mu_A(x_i)\rightarrow\mu_B(x_i)]\wedge[\mu_B(x_i)\rightarrow\mu_A(x_i)] + [\mu_{\bar{A}}(x_i)\rightarrow\mu_{\bar{B}}(x_i)]\wedge[\mu_{\bar{B}}(x_i)\rightarrow\mu_{\bar{A}}(x_i)]\}/2,$ $s_{L1}(A,B)=\sum_{i=1}^n[\mu_A(x_i)=\mu_B(x_i)]/n,$ $s_{L2}(A,B)=\sum_{i=1}^n[\mu_A(x_i)\Longleftrightarrow\mu_B(x_i)]/n$
Fuzzy-valued	
Dubois and Prade's method [10]	$s_F^\alpha(A,B)=\dfrac{\mathrm{card}((A_\alpha\cap B_\alpha)\cap\mathrm{supp}(A\cap B))}{\mathrm{card}((A_\alpha\cup B_\alpha)\cap\mathrm{supp}(A\cup B))}$, where supp means support

Note that those measures involving α-cuts require the FSs to be convex.

(4) Compute an IT2 FS similarity measure $s_M(\widetilde{A},\widetilde{B})$ as an average of T1 FS similarity measures s_{mn} that are computed for all of the MN combinations of the embedded T1 FSs for $\widetilde{A}$ and $\widetilde{B}$ (this uses the Representation Theorem in (A.6)), i.e.,

$$s_M(\widetilde{A},\widetilde{B})=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}s_{mn}, \tag{3}$$

where

$$s_{mn} = s(A_e^m, A_e^n) \quad (4)$$

and s_{mn} can be any T1 FS similarity measure, as in Table 1.

Mitchell's IT2 FS similarity measure has the following problems:

(1) Generally $s_M(\widetilde{A}, \widetilde{B})$ does not equal 1 even for the special case where $\widetilde{A}$ and $\widetilde{B}$ are exactly the same, because the randomly generated embedded T1 FSs from $\widetilde{A}$ and $\widetilde{B}$ will not always be the same.
(2) Because there are random numbers involved, $s_M(\widetilde{A}, \widetilde{B})$ may change from experiment[4] to experiment. When both M and N are large, some kind of stochastic convergence can be expected to occur (e.g., convergence in probability); however, the computational cost is heavy because the computation of (3) requires direct enumeration of all MN embedded T1 FSs.

3.2. Gorzalczany's IT2 FS compatibility measure

Gorzalczany proposed a compatibility measure for interval-valued FSs (IVFSs) [13]. Because an IVFS is an IT2 FS under a different name, the terms and symbols used in [13] are changed so that they are consistent with those in this paper.

Gorzalczany defined the *degree of compatibility*, $s_G(\widetilde{A}, \widetilde{B})$, between two IT2 FSs $\widetilde{A}$ and $\widetilde{B}$ as

$$s_G(\widetilde{A},\widetilde{B}) = \left[\min\left(\frac{\max_{x\in X}\{\min(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\underline{\mu}_{\widetilde{A}}(x)},\frac{\max_{x\in X}\{\min(\overline{\mu}_{\widetilde{A}}(x),\overline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\overline{\mu}_{\widetilde{A}}(x)}\right),\max\left(\frac{\max_{x\in X}\{\min(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\underline{\mu}_{\widetilde{A}}(x)},\frac{\max_{x\in X}\{\min(\overline{\mu}_{\widetilde{A}}(x),\overline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\overline{\mu}_{\widetilde{A}}(x)}\right)\right]. \quad (5)$$

Main properties of $s_G(\widetilde{A}, \widetilde{B})$ are: (1) $s_G(\widetilde{A}, \widetilde{A}) = [1, 1]$; (2) $s_G(\widetilde{A}, \widetilde{B}) = [0, 0]$ if and only if $\widetilde{A}$ and $\widetilde{B}$ are disjoint; and, (3) generally $s_G(\widetilde{A}, \widetilde{B}) \neq s_G(\widetilde{B}, \widetilde{A})$.

As pointed out by Tsiporkova and Zimmermann [40], compatibility measures do not "*perform consistently as similarity measures of FSs.*" It is easy to show that Gorzalczany's compatibility measure may give counter-intuitive results when used in linguistic approximation. Consider the example shown in Fig. 2, where $\max_{x\in X}\underline{\mu}_{\widetilde{A}}(x) = \max_{x\in X}\underline{\mu}_{\widetilde{B}}(x) = \mu_1$. Consequently,

$$\max_{x\in X}\{\min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\} = \max_{x\in X} \underline{\mu}_{\widetilde{A}}(x) = \mu_1 \quad (6)$$

and,

$$\frac{\max_{x\in X}\{\min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X} \underline{\mu}_{\widetilde{A}}(x)} = \frac{\mu_1}{\mu_1} = 1. \quad (7)$$

It is also easy to see that

$$\frac{\max_{x\in X}\{\min(\overline{\mu}_{\widetilde{A}}(x), \overline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X} \overline{\mu}_{\widetilde{A}}(x)} = \frac{1}{1} = 1. \quad (8)$$

Hence, for $\widetilde{A}$ and $\widetilde{B}$ shown in Fig. 2, $s_G(\widetilde{A}, \widetilde{B}) = [1, 1]$. Actually it can be shown that as long as $\max_{x\in X}\underline{\mu}_{\widetilde{A}}(x) = \max_{x\in X}\underline{\mu}_{\widetilde{B}}(x)$ and $\max_{x\in X}\overline{\mu}_{\widetilde{A}}(x) = \max_{x\in X}\overline{\mu}_{\widetilde{B}}(x)$, no matter how different the shapes of $\widetilde{A}$ and $\widetilde{B}$ are, Gorzalczany's compatibility measure always gives $s_G(\widetilde{A}, \widetilde{B}) = s_G(\widetilde{B}, \widetilde{A}) = [1, 1]$, which is counter-intuitive.

[4] One experiment is comprised of M (N) randomly chosen embedded T1 FSs for $\widetilde{A}$ ($\widetilde{B}$).

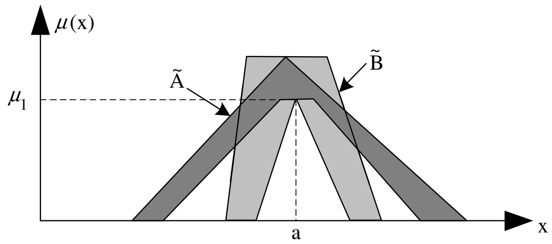

Fig. 2. Example for Gorzalczany's compatibility measure, which gives $s_G(\widetilde{A},\widetilde{B}) = [1,1]$.

3.3. Bustince's IT2 FS similarity measure

Bustince also proposed a similarity measure for IVFSs [6]. Again, the terms and symbols used in [6] are changed so that they are consistent with those in this paper.

First, Bustince defined a *normal interval valued similarity measure* $s_B(\widetilde{A},\widetilde{B})$ between two IT2 FSs $\widetilde{A}$ and $\widetilde{B}$, as one that satisfies the following five properties: (1) $s_B(\widetilde{A},\widetilde{B}) = s_B(\widetilde{B},\widetilde{A})$; (2) for a crisp set a and its *complement* $c(a)$, $s_B(a,c(a)) = 0$; (3) $s_B(\widetilde{A},\widetilde{A}) = [1,1]$; (4) if $\widetilde{A} \leqslant \widetilde{B} \leqslant \widetilde{C}$, then $s_B(\widetilde{A},\widetilde{B}) \geqslant s_B(\widetilde{A},\widetilde{C})$ and $s_B(\widetilde{B},\widetilde{C}) \geqslant s_B(\widetilde{A},\widetilde{C})$; and, (5) if A and B are T1 FSs, then $s_B(A,B) \in [0,1]$, i.e. $s_B(A,B)$ reduces to a number. Note that $\widetilde{A} \leqslant \widetilde{B}$ if and only if for $\forall x \in X$, $\underline{\mu}_{\widetilde{A}}(x) \leqslant \underline{\mu}_{\widetilde{B}}(x)$ and $\overline{\mu}_{\widetilde{A}}(x) \leqslant \overline{\mu}_{\widetilde{B}}(x)$.

He then proposed

$$s_B(\widetilde{A},\widetilde{B}) = [s_L(\widetilde{A},\widetilde{B}), s_U(\widetilde{A},\widetilde{B})] \tag{9}$$

as an *interval-valued normal similarity measure*, where

$$s_L(\widetilde{A},\widetilde{B}) = \Upsilon_L(\widetilde{A},\widetilde{B}) \bigstar \Upsilon_L(\widetilde{B},\widetilde{A}) \tag{10}$$

and

$$s_U(\widetilde{A},\widetilde{B}) = \Upsilon_U(\widetilde{A},\widetilde{B}) \bigstar \Upsilon_U(\widetilde{B},\widetilde{A}), \tag{11}$$

$\bigstar$ can be any *t*-norm (e.g., minimum), and $[\Upsilon_L(\widetilde{A},\widetilde{B}), \Upsilon_U(\widetilde{A},\widetilde{B})]$ is an *interval valued inclusion grade indicator* [6] of $\widetilde{A}$ in $\widetilde{B}$. $\Upsilon_L(\widetilde{A},\widetilde{B})$ and $\Upsilon_U(\widetilde{A},\widetilde{B})$ used in this paper (and taken from [6]) are computed as

$$\Upsilon_L(\widetilde{A},\widetilde{B}) = \inf_{x\in X}\{1, \min(1 - \underline{\mu}_{\widetilde{A}}(x) + \underline{\mu}_{\widetilde{B}}(x), 1 - \overline{\mu}_{\widetilde{A}}(x) + \overline{\mu}_{\widetilde{B}}(x))\} \tag{12}$$

$$\Upsilon_U(\widetilde{A},\widetilde{B}) = \inf_{x\in X}\{1, \max(1 - \underline{\mu}_{\widetilde{A}}(x) + \underline{\mu}_{\widetilde{B}}(x), 1 - \overline{\mu}_{\widetilde{A}}(x) + \overline{\mu}_{\widetilde{B}}(x))\} \tag{13}$$

Both Bustince's and Gorzalczany's similarity measures, $s_B(\widetilde{A},\widetilde{B})$ and $s_G(\widetilde{A},\widetilde{B})$, are intervals, and $s_B(\widetilde{A},\widetilde{B})$ has the desirable property that $s_B(\widetilde{A},\widetilde{B}) = s_B(\widetilde{B},\widetilde{A})$ whereas $s_G(\widetilde{A},\widetilde{B})$ does not.

Major differences between Bustince's and Mitchell's similarity measures, $s_B(\widetilde{A},\widetilde{B})$ and $s_M(\widetilde{A},\widetilde{B})$, are: (1) the former chooses $s_B(\widetilde{A},\widetilde{B})$ to satisfy a set of five similarity-measure properties whereas the latter does not; and, (2) $s_B(\widetilde{A},\widetilde{B})$ is an interval whereas $s_M(\widetilde{A},\widetilde{B})$ is a point-value.

A problem with Bustince's similarity measure is that when $\widetilde{A}$ and $\widetilde{B}$ are disjoint, no matter how far away they are from each other, $s_B(\widetilde{A},\widetilde{B})$ will always be the same. For a simple example to demonstrate this, consider the case where disjoint $\widetilde{A}$ and $\widetilde{B}$ have exactly the same shape, as shown in Fig. 3a. In this case, $1 - \underline{\mu}_{\widetilde{A}}(x) + \underline{\mu}_{\widetilde{B}}(x)$ and $1 - \overline{\mu}_{\widetilde{A}}(x) + \overline{\mu}_{\widetilde{B}}(x)$ are shown in Fig. 3b as the dashed lines and the solid lines, respectively, and, $\min(1 - \underline{\mu}_{\widetilde{A}}(x) + \underline{\mu}_{\widetilde{B}}(x), 1 - \overline{\mu}_{\widetilde{A}}(x) + \overline{\mu}_{\widetilde{B}}(x))$ and $\max(1 - \underline{\mu}_{\widetilde{A}}(x) + \underline{\mu}_{\widetilde{B}}(x), 1 - \overline{\mu}_{\widetilde{A}}(x) + \overline{\mu}_{\widetilde{B}}(x))$ are shown in Fig. 3c as the dashed lines and solid lines, respectively. Substituting the two functions in Fig. 3c into (12) and (13), observe that $\Upsilon_L(\widetilde{A},\widetilde{B}) = 0$ and $\Upsilon_U(\widetilde{A},\widetilde{B}) = 1 - \mu_1$ [indicated by a square in Fig. 3c]. In a similar way (see Figs. 3d and e), it is easy to show $\Upsilon_L(\widetilde{B},\widetilde{A}) = 0$ and $\Upsilon_U(\widetilde{B},\widetilde{A}) = 1 - \mu_1$. Consequently, in (10) $s_L(\widetilde{A},\widetilde{B}) = 0\bigstar 0$ and in (11) $s_U(\widetilde{A},\widetilde{B}) = (1-\mu_1)\bigstar(1-\mu_1)$ so that $s_B(\widetilde{A},\widetilde{B}) = [0\bigstar 0, (1-\mu_1)\bigstar(1-\mu_1)]$. As long as $\widetilde{A}$ and $\widetilde{B}$ are disjoint, i.e. $d \geqslant 0$ in Fig. 3a, $s_B(\widetilde{A},\widetilde{B})$ is always $[0\bigstar 0, (1-\mu_1)\bigstar(1-\mu_1)]$ regardless of d, and usually $(1-\mu_1)\bigstar(1-\mu_1) \neq 0$. When $\widetilde{A}$ and $\widetilde{B}$ are disjoint, $s_B(\widetilde{A},\widetilde{B})$ is expected to either decrease as d increases or be 0; hence, Bustince's similarity measure is counter-intuitive for this situation.

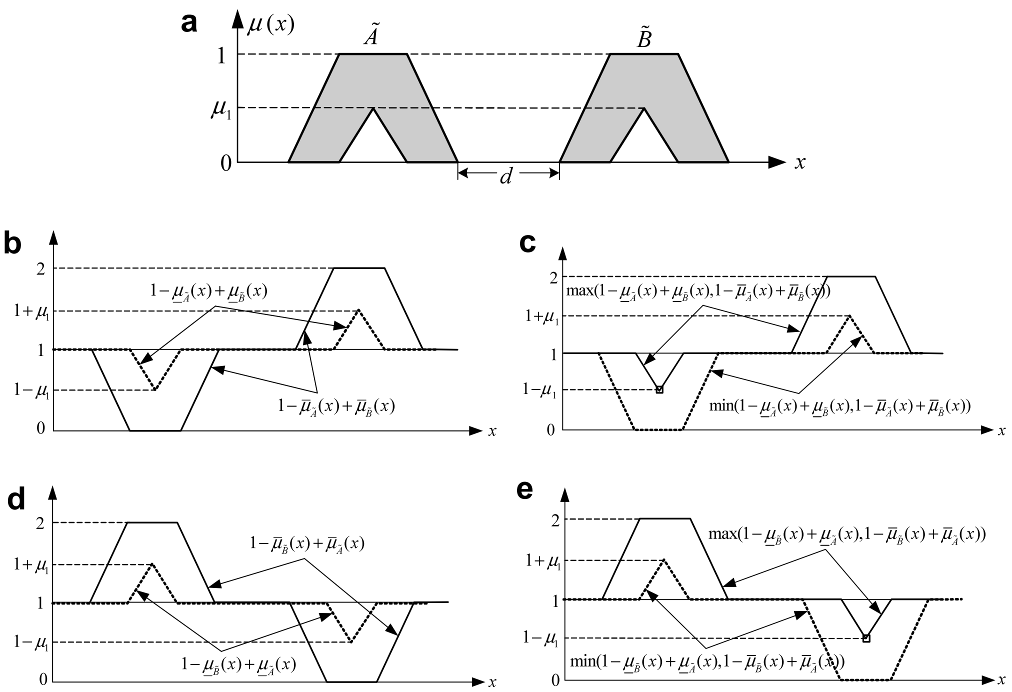

Fig. 3. Example for Bustince's similarity measure when $\widetilde{A}$ and $\widetilde{B}$ are disjoint. (a) $\widetilde{A}$ and $\widetilde{B}$; (b) and (c) functions needed to compute $\Upsilon_{\mathrm{L}}(\widetilde{A},\widetilde{B})$ and $\Upsilon_{\mathrm{U}}(\widetilde{A},\widetilde{B})$ in (12) and (13); and, (d) and (e) functions needed to compute $\Upsilon_{\mathrm{L}}(\widetilde{B},\widetilde{A})$ and $\Upsilon_{\mathrm{U}}(\widetilde{B},\widetilde{A})$ in (12) and (13).

3.4. Zeng and Li's IT2 FS similarity measure

Zeng and Li's similarity measure was also proposed for IVFSs [51]. Again, the terms and symbols used in [51] are changed so that they are consistent with those in this paper.

Zeng and Li defined a real function s_{Z}: $\widetilde{A}\times\widetilde{B}\to[0,1]$ as a similarity measure of IT2 FSs, if s_{Z} satisfies the following properties: (1) $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})=s_{\mathrm{Z}}(\widetilde{B},\widetilde{A})$; (2) for a crisp set a and its complement $c(a)$, $s_{\mathrm{Z}}(a,c(a))=0$; (3) $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})=1\iff\widetilde{A}=\widetilde{B}$; and, (4) if $\widetilde{A}\leqslant\widetilde{B}\leqslant\widetilde{C}$, then $s_{\mathrm{Z}}(\widetilde{A},\widetilde{C})\leqslant s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})$ and $s_{\mathrm{Z}}(\widetilde{A},\widetilde{C})\leqslant s_{\mathrm{Z}}(\widetilde{B},\widetilde{C})$.

They then proposed the following similarity measure for IT2 FSs if the universes of discourse of $\widetilde{A}$ and $\widetilde{B}$ are discrete:

$$s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})=1-\frac{1}{2n}\sum_{i=1}^{n}(|\underline{\mu}_{\tilde{A}}(x_i)-\underline{\mu}_{\tilde{B}}(x_i)|+|\overline{\mu}_{\tilde{A}}(x_i)-\overline{\mu}_{\tilde{B}}(x_i)|), \tag{14}$$

and, if the universes of discourse of $\widetilde{A}$ and $\widetilde{B}$ are continuous in $[a,b]$,

$$s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})=1-\frac{1}{2(b-a)}\int_a^b(|\underline{\mu}_{\tilde{A}}(x)-\underline{\mu}_{\tilde{B}}(x)|+|\overline{\mu}_{\tilde{A}}(x)-\overline{\mu}_{\tilde{B}}(x)|)\mathrm{d}x. \tag{15}$$

Properties of Zeng and Li's similarity measure are quite similar to those of Bustince's. The main difference is that the former treats the similarity measure as a crisp number, whereas the latter gives an interval. Zeng and Li's similarity measure has a problem similar to that of Bustince's, but may be worse depending on the choice of a and b (see (15)). For example in Fig. 4, $\widetilde{B}$ and $\widetilde{B}'$ have the same shape but are at different distances from $\widetilde{A}$; hence, $\int_a^b(|\underline{\mu}_{\tilde{A}}(x)-\underline{\mu}_{\tilde{B}}(x)|+|\overline{\mu}_{\tilde{A}}(x)-\overline{\mu}_{\tilde{B}}(x)|)\mathrm{d}x$, $\int_a^{b'}(|\underline{\mu}_{\tilde{A}}(x)-\underline{\mu}_{\tilde{B}}(x)|+|\overline{\mu}_{\tilde{A}}(x)-\overline{\mu}_{\tilde{B}}(x)|)\mathrm{d}x$ and $\int_a^{b'}(|\underline{\mu}_{\tilde{A}}(x)-\underline{\mu}_{\tilde{B}'}(x)|+|\overline{\mu}_{\tilde{A}}(x)-\overline{\mu}_{\tilde{B}'}(x)|)\mathrm{d}x$ are equal, and this value is denoted as c. There can be two methods in computing $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})$ and $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B}')$:

(1) If the interval $[a,b]$ is used to compute $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})$ and $[a,b']$ is used to compute $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B}')$, then $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})=1-c/[2(b-a)]$ and $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B}')=1-c/[2(b'-a)]$. Because $b'-a>b-a$, $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B})<s_{\mathrm{Z}}(\widetilde{A},\widetilde{B}')$, which means $\widetilde{B}'$ is more similar to $\widetilde{A}$ than $\widetilde{B}$ is. Additionally, as $b'-a$ increases, $s_{\mathrm{Z}}(\widetilde{A},\widetilde{B}')$ approaches 1.

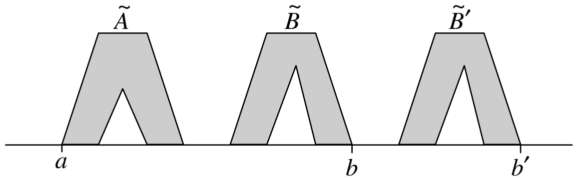

Fig. 4. An example of Zeng and Li's similarity measure for disjoint IT2 FSs.

(2) If the interval $[a, b']$ is used to compute both $s_Z(\widetilde{A}, \widetilde{B})$ and $s_Z(\widetilde{A}, \widetilde{B}')$, then $s_Z(\widetilde{A}, \widetilde{B}) = 1 - c/[2(b' - a)]$ and $s_Z(\widetilde{A}, \widetilde{B}') = 1 - c/[2(b' - a)]$; hence, $s_Z(\widetilde{A}, \widetilde{B}) = s_Z(\widetilde{A}, \widetilde{B}') > 0$.

Both methods produce results that are counter-intuitive, because (Fig. 4) we should have $s(\widetilde{A}, \widetilde{B}) > s(\widetilde{A}, \widetilde{B}')$. If this is not true, another reasonable result is $s(\widetilde{A}, \widetilde{B}) = s(\widetilde{A}, \widetilde{B}') = 0$, instead of a non-zero constant as given by Method (2).

3.5. Summary

A summary of all similarity/compatibility measures for IT2 FSs introduced in this section is given in Table 2. It is worth noting that each of the four existing similarity/compatibility measures for IT2 FSs has its problems:

- Mitchell's similarity measure involves randomness which can lead to different answers, and the computational cost is high.
- Gorzalczany's, Bustince's, and Zeng and Li's similarity/compatibility measures give counter-intuitive results for some special cases.

Table 2
Summary of existing similarity/compatibility measures for IT2 FSs

Measure	Equation
Mitchell's method [34]	$s_M(\widetilde{A}, \widetilde{B}) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} s_{mn}$, where $s_{mn} = s(A_e^m, A_e^n)$, and s can be any similarity measure for T1 FSs
Gorzalczany's method [13]	$s_G(\widetilde{A}, \widetilde{B}) = \left[\min\left(\frac{\max_{x\in X}\{\min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\underline{\mu}_{\widetilde{A}}(x)}, \frac{\max_{x\in X}\{\min(\overline{\mu}_{\widetilde{A}}(x), \overline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\overline{\mu}_{\widetilde{A}}(x)}\right), \max\left(\frac{\max_{x\in X}\{\min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\underline{\mu}_{\widetilde{A}}(x)}, \frac{\max_{x\in X}\{\min(\overline{\mu}_{\widetilde{A}}(x), \overline{\mu}_{\widetilde{B}}(x))\}}{\max_{x\in X}\overline{\mu}_{\widetilde{A}}(x)}\right)\right]$
Bustince's method [6]	$s_B(\widetilde{A}, \widetilde{B}) = [s_L(\widetilde{A}, \widetilde{B}), s_U(\widetilde{A}, \widetilde{B})]$, where $s_L(\widetilde{A}, \widetilde{B}) = \Upsilon_L(\widetilde{A}, \widetilde{B}) \bigstar \Upsilon_L(\widetilde{B}, \widetilde{A})$, $s_U(\widetilde{A}, \widetilde{B}) = \Upsilon_U(\widetilde{A}, \widetilde{B}) \bigstar \Upsilon_U(\widetilde{B}, \widetilde{A})$ Examples of Υ_L and Υ_U are given in (12) and (13).
Zeng and Li's method [51]	For discrete universe of discourse, $s_Z(\widetilde{A}, \widetilde{B}) = 1 - \frac{1}{2n}\sum_{i=1}^{n}(\lvert\underline{\mu}_{\widetilde{A}}(x_i) - \underline{\mu}_{\widetilde{B}}(x_i)\rvert + \lvert\overline{\mu}_{\widetilde{A}}(x_i) - \overline{\mu}_{\widetilde{B}}(x_i)\rvert)$ For continuous universe of discourse, $s_Z(\widetilde{A}, \widetilde{B}) = 1 - \frac{1}{2(b-a)}\int_a^b(\lvert\underline{\mu}_{\widetilde{A}}(x) - \underline{\mu}_{\widetilde{B}}(x)\rvert + \lvert\overline{\mu}_{\widetilde{A}}(x) - \overline{\mu}_{\widetilde{B}}(x)\rvert)\mathrm{d}x$

3.6. Proposed properties for a similarity measure

To avoid the problems just mentioned, the following four properties are proposed for a similarity measure for IT2 FSs.

(P.1) *Reflexivity*: The similarity between two IT2 FSs is 1 if and only if they are exactly the same, i.e. $s(\widetilde{A},\widetilde{B}) = 1 \Longleftrightarrow \widetilde{A} = \widetilde{B}$.
(P.2) *Symmetry*: The similarity between two IT2 FSs should be a constant regardless of the order in which they are compared, i.e. $s(\widetilde{A},\widetilde{B}) = s(\widetilde{B},\widetilde{A})$.
(P.3) *Transitivity*: (a) If three IT2 FSs have the same shape, then the similarity between two nearby IT2 FSs should be larger than the similarity between two further away IT2 FSs, i.e. if $\widetilde{A}$, $\widetilde{B}$ and $\widetilde{C}$ are of the same shape and $c(\widetilde{A}) < c(\widetilde{B}) < c(\widetilde{C})$ (see the definition of $c(\widetilde{A})$ in (17)), then $s(\widetilde{A},\widetilde{B}) > s(\widetilde{A},\widetilde{C})$ and $s(\widetilde{B},\widetilde{C}) > s(\widetilde{A},\widetilde{C})$; (b) If $c(\widetilde{A}) = c(\widetilde{B}) = c(\widetilde{C})$ and $\widetilde{A} < \widetilde{B} < \widetilde{C}$, then $s(\widetilde{A},\widetilde{B}) > s(\widetilde{A},\widetilde{C})$ and $s(\widetilde{B},\widetilde{C}) > s(\widetilde{A},\widetilde{C})$.
(P.4) *Overlap*: If two IT2 FSs partially overlap, then there should be some similarity between them, i.e. if there exists at least one x with non-zero memberships on both $\widetilde{A}$ and $\widetilde{B}$, then $s(\widetilde{A},\widetilde{B}) > 0$.

In the next section a vector similarity measure is proposed which possesses these properties.

4. A vector similarity measure for IT2 FSs

Recall that our goal is to find the B_i which most closely resembles $\widetilde{A}$; hence, when the similarity of two IT2 FSs $\widetilde{A}$ and $\widetilde{B}$ are compared, it is necessary to compare their shapes as well as proximity. In this section, a vector similarity measure (VSM) for IT2 FSs, $\mathbf{s}_v(\widetilde{A},\widetilde{B})$, is proposed, one that has two components, i.e.,

$$\mathbf{s}_v(\widetilde{A},\widetilde{B}) = (s_1(\widetilde{A},\widetilde{B}), s_2(\widetilde{A},\widetilde{B}))^{\mathrm{T}} \tag{16}$$

where $s_1(\widetilde{A},\widetilde{B}) \in [0,1]$ is a similarity measure on the shapes of $\widetilde{A}$ and $\widetilde{B}$, and $s_2(\widetilde{A},\widetilde{B}) \in [0,1]$ is a similarity measure on the proximity of $\widetilde{A}$ and $\widetilde{B}$.

4.1. Definition of $s_1(\widetilde{A},\widetilde{B})$

Because the proximity of $\widetilde{A}$ and $\widetilde{B}$ is considered in $s_2(\widetilde{A},\widetilde{B})$, in computing $s_1(\widetilde{A},\widetilde{B})$ $\widetilde{A}$ and $\widetilde{B}$ are "aligned" so that their shapes can be compared. Denote the centroids (see Appendix A) of $\widetilde{A}$ and $\widetilde{B}$ as $C_{\tilde{A}} = [c_l(\widetilde{A}), c_r(\widetilde{A})]$ and $C_{\tilde{B}} = [c_l(\widetilde{B}), c_r(\widetilde{B})]$, respectively, and the centers of $C_{\tilde{A}}$ and $C_{\tilde{B}}$ as

$$c(\widetilde{A}) = [c_l(\widetilde{A}) + c_r(\widetilde{A})]/2 \tag{17}$$

$$c(\widetilde{B}) = [c_l(\widetilde{B}) + c_r(\widetilde{B})]/2 \tag{18}$$

A reasonable alignment method is to move one or both of $\widetilde{A}$ and $\widetilde{B}$ so that $c(\widetilde{A})$ and $c(\widetilde{B})$ coincide (see Fig. 5). The two IT2 FSs can be moved to any location as long as $c(\widetilde{A})$ and $c(\widetilde{B})$ coincide; this will not affect the value of $s_1(\widetilde{A},\widetilde{B})$. In this paper $\widetilde{B}$ is moved to $\widetilde{A}$ and called $\widetilde{B}'$, as shown in Fig. 5b.

$s_1(\widetilde{A},\widetilde{B})$ may be defined as a crisp number, or an interval; however, as shown next, there may be problems when defining $s_1(\widetilde{A},\widetilde{B})$ as an interval.

An intuitive interval realization of $s_1(\widetilde{A},\widetilde{B})$ is to define it as an extension of Jaccard's similarity measure (see Table 1; note that f is chosen to be the cardinality in this paper) from T1 FSs to IT2 FSs by using the Representation Theorem (Appendix A.1), i.e.

$$s_{1,\text{interval}}(\widetilde{A},\widetilde{B}) = \bigcup_{\forall A_e, B'_e} \frac{\text{card}(A_e \cap B'_e)}{\text{card}(A_e \cup B'_e)} = [s_{1,l}, s_{1,r}] \tag{19}$$

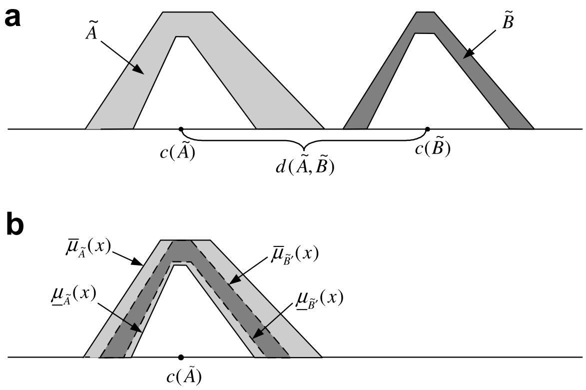

Fig. 5. An example for the proposed VSM. (a) $c(\tilde{A})$ and $c(\tilde{B})$ denote the center of the centroids of $\tilde{A}$ and $\tilde{B}$, respectively; (b) $\tilde{B}'$ is obtained by moving $\tilde{B}$ so that $c(\tilde{B})$ coincides with $c(\tilde{A})$. The solid curves are for $\tilde{A}$ and the dashed curves are for $\tilde{B}'$.

where A_e and B'_e are embedded T1 FSs of $\tilde{A}$ and $\tilde{B}'$, respectively, and

$$s_{1,l} \equiv \min_{\forall A_e, B'_e} \frac{\mathrm{card}(A_e \cap B'_e)}{\mathrm{card}(A_e \cup B'_e)} \tag{20}$$

$$s_{1,r} \equiv \max_{\forall A_e, B'_e} \frac{\mathrm{card}(A_e \cap B'_e)}{\mathrm{card}(A_e \cup B'_e)} \tag{21}$$

Unfortunately, there are no closed-form solutions for $s_{1,l}$ and $s_{1,r}$. Furthermore, even if we can compute $s_{1,l}$ and $s_{1,r}$, there is still a need to convert $[s_{1,l}, s_{1,r}]$ to a crisp number because in many applications ranking of similarities are needed.

For simplicity, in this paper $s_1(\tilde{A},\tilde{B})$ is defined as a crisp number equal to the ratio of the *average cardinalities* (see (A.13)) of[5] FOU$(\tilde{A}\cap\tilde{B}')$ and[6] FOU$(\tilde{A}\cup\tilde{B}')$, i.e.

$$s_1(\tilde{A},\tilde{B}) \equiv \frac{AC[\mathrm{FOU}(\tilde{A}\cap\tilde{B}')]}{AC[\mathrm{FOU}(\tilde{A}\cup\tilde{B}')]} \tag{22}$$

$$= \frac{\mathrm{card}(\bar{\mu}_{\tilde{A}}(x)\cap\bar{\mu}_{\tilde{B}'}(x)) + \mathrm{card}(\underline{\mu}_{\tilde{A}}(x)\cap\underline{\mu}_{\tilde{B}'}(x))}{\mathrm{card}(\bar{\mu}_{\tilde{A}}(x)\cup\bar{\mu}_{\tilde{B}'}(x)) + \mathrm{card}(\underline{\mu}_{\tilde{A}}(x)\cup\underline{\mu}_{\tilde{B}'}(x))} \tag{23}$$

$$= \frac{\int_X \min(\bar{\mu}_{\tilde{A}}(x),\bar{\mu}_{\tilde{B}'}(x))\mathrm{d}x + \int_X \min(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x}{\int_X \max(\bar{\mu}_{\tilde{A}}(x),\bar{\mu}_{\tilde{B}'}(x))\mathrm{d}x + \int_X \max(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x}, \tag{24}$$

where $\bar{\mu}_{\tilde{B}'}(x)$ and $\underline{\mu}_{\tilde{B}'}(x)$ are illustrated in Fig. 5b. When all uncertainty disappears, $\tilde{A}$ and $\tilde{B}$ become T1 FSs A and B, and (24) reduces to Jaccard's similarity measure (see Table 1, in which f is chosen as the cardinality).

$s_1(\tilde{A},\tilde{B})$ has the following properties:

Theorem 1. *(a)* $0 \leqslant s_1(\tilde{A},\tilde{B}) \leqslant 1$; *(b)* $s_1(\tilde{A},\tilde{B}) = 1 \Longleftrightarrow \tilde{A} = \tilde{B}'$, *i.e.* $\tilde{A}$ *and* $\tilde{B}$ *have the same shape*; *and*, *(c)* $s_1(\tilde{A},\tilde{B}) = s_1(\tilde{B},\tilde{A})$.

Proof. See Appendix B.1. □

4.2. Definition of $s_2(\tilde{A},\tilde{B})$

$s_2(\tilde{A},\tilde{B})$ measures the proximity of $\tilde{A}$ and $\tilde{B}$, and is defined as

$$s_2(\tilde{A},\tilde{B}) \equiv h(d(\tilde{A},\tilde{B})) \tag{25}$$

[5] $\tilde{A}\bigcap\tilde{B}' = 1/\bigcup_{\forall x\in X}[\underline{\mu}_{\tilde{A}}(x)\bigstar\underline{\mu}_{\tilde{B}'}(x), \bar{\mu}_{\tilde{A}}(x)\bigstar\bar{\mu}_{\tilde{B}'}(x)]$, where $\bigstar$ is a t-norm. In (24) the min t-norm is used [25].

[6] $\tilde{A}\bigcup\tilde{B}' = 1/\bigcup_{\forall x\in X}[\underline{\mu}_{\tilde{A}}(x)\vee\underline{\mu}_{\tilde{B}'}(x), \bar{\mu}_{\tilde{A}}(x)\vee\bar{\mu}_{\tilde{B}'}(x)]$, where $\vee$ is a t-conorm. In (24) the max t-conorm is used [25].

where $d(\widetilde{A},\widetilde{B}) = |c(\widetilde{A}) - c(\widetilde{B})|$ is the Euclidean distance between the centers of the centroids of $\widetilde{A}$ and $\widetilde{B}$ (see Fig. 5a), and h can be any function satisfying: (1) $\lim_{x\to\infty} h(x) = 0$; (2) $h(x) = 1$ if and only if $x = 0$; and, (3) $h(x)$ decreases monotonically as x increases.

Theorem 2. *$s_2(\widetilde{A},\widetilde{B}) \in [0,1]$, and $s_2(\widetilde{A},\widetilde{B}) = 1$ if and only if $c(\widetilde{A}) = c(\widetilde{B})$.*

Proof. Theorem 2 is obvious from (25) and the above constraints on $h(x)$. □

An example of $s_2(\widetilde{A},\widetilde{B})$ is

$$s_2(\widetilde{A},\widetilde{B}) = e^{-rd(\widetilde{A},\widetilde{B})}, \tag{26}$$

where r is a positive constant. $s_2(\widetilde{A},\widetilde{B})$ is chosen as an exponential function because we believe the similarity between two FSs should decrease rapidly as the distance between them increases.

4.3. On converting $\mathbf{s}_v(\widetilde{A},\widetilde{B})$ to the scalar similarity measure $s_s(\widetilde{A},\widetilde{B})$

$\mathbf{s}_v(\widetilde{A},\widetilde{B})$ enables us to separately quantify the similarity of two features, shape and proximity. As mentioned in Section 1, in CWW $\mathbf{s}_v(\widetilde{A},\widetilde{B}_i)$ $(i = 1,2,\ldots,N)$ need to be ranked to find the $\widetilde{B}_i$ most similar to A. This can be achieved by first converting the vector $\mathbf{s}_v(\widetilde{A},\widetilde{B}_i)$ to a scalar similarity measure $s_s(\widetilde{A},\widetilde{B}_i)$ and then ranking $s_s(\widetilde{A},\widetilde{B}_i)$ $(i = 1,2,\ldots,N)$.

In this paper, the scalar similarity between two IT2 FSs $\widetilde{A}$ and $\widetilde{B}$ is computed as the product of their similarities in shape and proximity,[7] i.e.

$$s_s(\widetilde{A},\widetilde{B}) = s_1(\widetilde{A},\widetilde{B}) \times s_2(\widetilde{A},\widetilde{B}) \tag{27}$$

Properties of $s_s(\widetilde{A},\widetilde{B})$ include:

Theorem 3. *(a) $\widetilde{A} = \widetilde{B} \Longleftrightarrow s_s(\widetilde{A},\widetilde{B}) = 1$; (b) $s_s(\widetilde{A},\widetilde{B}) = s_s(\widetilde{B},\widetilde{A})$; (c) $s_s(\widetilde{A},\widetilde{B}) > s_s(\widetilde{A},\widetilde{C})$ and $s_s(\widetilde{B},\widetilde{C}) > s_s(\widetilde{A},\widetilde{C})$ if $\widetilde{A}$, $\widetilde{B}$ and $\widetilde{C}$ have the same shape and $c(\widetilde{A}) < c(\widetilde{B}) < c(\widetilde{C})$; (d) $s_s(\widetilde{A},\widetilde{B}) > s_s(\widetilde{A},\widetilde{C})$ and $s_s(\widetilde{B},\widetilde{C}) > s_s(\widetilde{A},\widetilde{C})$ if $c(\widetilde{A}) = c(\widetilde{B}) = c(\widetilde{C})$ and $\widetilde{A} < \widetilde{B} < \widetilde{C}$; and, (e) $s_s(\widetilde{A},\widetilde{B}) > 0$.*

Proof. See Appendix B.2. □

Theorem 3 shows that $s_s(\widetilde{A},\widetilde{B})$ satisfies the four properties stated in Section 3.6.

4.4. Example

Comparisons of all the similarity measures for IT2 FSs introduced in this paper are given in Table 3 for IT2 FSs $\widetilde{A} - \widetilde{G}$ depicted in Fig. 6. Note that $\widetilde{A} - \widetilde{E}$ have the same shape. The domain of x (e.g., the support of $\widetilde{A} \cup \widetilde{B}$ in computing $s(\widetilde{A},\widetilde{B})$) was discretized into 500 equal-length intervals, $M \equiv N = 10$ in Mitchell's similarity measure, Method (1) in Section 3.4 was used to choose x_i in Zeng and Li's similarity measure, and $r \equiv 4/|X|$ ($|X|$ is the length of the support of $\widetilde{A} \cup \widetilde{B}$) in the VSM (see (26)).

Observe from Table 3 that the outputs of the VSM are reasonable for all six cases, according to the four properties proposed in Section 3.6. Observe also that:

(1) Using Mitchell's method, $s(\widetilde{F},\widetilde{F}) = 0.6007$, which should be 1.
(2) Using Gorzalczany's method, $s(\widetilde{F},\widetilde{G}) = [1,1]$, which should be less than 1.
(3) Using Bustince's method, $s(\widetilde{A},\widetilde{D}) = s(\widetilde{A},\widetilde{E}) \neq [0,0]$, which should be $s(\widetilde{A},\widetilde{D}) > s(\widetilde{A},\widetilde{E})$, or at least, $s(\widetilde{A},\widetilde{D}) = s(\widetilde{A},\widetilde{E}) = [0,0]$. Besides, $s(\widetilde{A},\widetilde{B})$, $s(\widetilde{A},\widetilde{C})$ and $s(\widetilde{A},\widetilde{D})$ are difficult to distinguish.
(4) Using Zeng and Li's method, $s(\widetilde{A},\widetilde{B}) < s(\widetilde{A},\widetilde{D}) < s(\widetilde{A},\widetilde{E})$, which should be $s(\widetilde{A},\widetilde{B}) > s(\widetilde{A},\widetilde{D}) > s(\widetilde{A},\widetilde{E})$, and, $s(\widetilde{A},\widetilde{C}) < s(\widetilde{A},\widetilde{D}) < s(\widetilde{A},\widetilde{E})$, which should be $s(\widetilde{A},\widetilde{C}) > s(\widetilde{A},\widetilde{D}) > s(\widetilde{A},\widetilde{E})$.

[7] Recently, Bonissone et al. [5] defined a similarity measure as a weighted minimum of several sub-similarity measures. Although similar to our idea, their objective is quite different from our objective; hence, their similarity measure is not used in this paper.

Table 3
Comparison of similarity measures for IT2 FSs $\widetilde{A}-\widetilde{G}$ shown in Fig. 6

Similarity measure	$s(\widetilde{A},\widetilde{B})$	$s(\widetilde{A},\widetilde{C})$	$s(\widetilde{A},\widetilde{D})$	$s(\widetilde{A},\widetilde{E})$	$s(\widetilde{F},\widetilde{F})$	$s(\widetilde{F},\widetilde{G})$
Mitchell's method (s_M)	0.1494	0.0124	0	0	0.6007	0.5762
Gorzalczany's method (s_G)	[0, 0.5980]	[0, 0.1967]	[0, 0]	[0, 0]	[1, 1]	[1, 1]
Bustince's method (s_B)	[0.0017, 0.2016]	[0.0010, 0.2016]	[0, 0.2016]	[0, 0.2016]	[1, 1]	[0.3337, 1]
Zeng and Li's method (s_Z)	0.6578	0.6452	0.7006	0.7467	1	0.7782
VSM (s_s)	0.2013	0.0406	0.0082	0.0017	1	0.5732

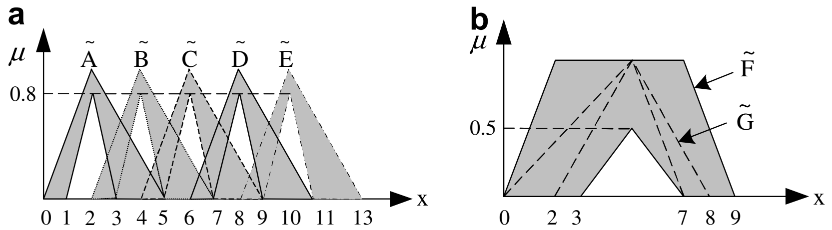

Fig. 6. Examples used in the comparative study: (a) $\widetilde{A}-\widetilde{E}$, which have the same shape; (b) $\widetilde{F}$ (solid lines) and $\widetilde{G}$ (dashed lines).

Finally, observe that the VSM does not have any of the short-comings of these four similarity measures. This example demonstrates that our VSM has the potential to succeed when it is used as the decoder in the Per-C shown in Fig. 1.

5. Discussions

5.1. More about an interval VSM for IT2 FSs

Scalar similarity measures are used for T1 FSs, so, it may be more reasonable to use interval similarity measures for IT2 FSs, as Gorzalczany and Bustince have done, because an IT2 FS has an extra degree of freedom than a T1 FS. However, recall that the objective of this paper is to identify an IT2 FS, $\widetilde{B}_k$, from a group of IT2 FSs, $\widetilde{B}_i$ $(i = 1, \ldots, N)$, so that $\widetilde{B}_k$ is most similar to a target IT2 FS, $\widetilde{A}$. Consequently, a crisp similarity measure is needed so that $s(\widetilde{A},\widetilde{B}_i)$ can be ranked. Even if an interval VSM for IT2 FSs was developed, its outputs would have to be converted to scalars before a ranking could be made. This may increase the computational cost. Besides, simulation results have shown that the output of $\mathbf{s}_v(\widetilde{A},\widetilde{B})$ is reasonable. So, it is unnecessary to develop an interval VSM for our application.

5.2. The VSM for T1 FSs

Because T1 FSs are special cases of IT2 FSs when all uncertainty disappears, the VSM for IT2 FSs developed in Section 4 can also be used for T1 FSs, as shown in this subsection.

When $\widetilde{A}$ and $\widetilde{B}$ reduce to T1 FSs, A and B, $\mathbf{s}_v(\widetilde{A},\widetilde{B})$ becomes a VSM for T1 FSs, $\mathbf{s}_v(A,B)$, i.e.,

$$\mathbf{s}_v(A,B) = (s_1(A,B), s_2(A,B))^{\mathrm{T}}, \tag{28}$$

where $s_1(A,B) \in [0,1]$ is a similarity measure on the shapes of A and B, and $s_2(A,B) \in [0,1]$ is a similarity measure on the proximity of A and B. Again, to define $\mathbf{s}_v(A,B)$, $s_1(A,B)$ and $s_2(A,B)$ must first be defined.

5.2.1. Definition of $s_1(A,B)$

Because the proximity of A and B is considered in $s_2(A,B)$, when computing $s_1(A,B)$ A and B are also "aligned" so that their shapes can be compared. In the IT2 FSs case one or both of $\widetilde{A}$ and $\widetilde{B}$ are moved so that $c(\widetilde{A})$ coincided with $c(\widetilde{B})$. When A and B are T1 FSs, one or both of A and B are moved so that their centroids $c(A)$ and $c(B)$ coincide. In this paper B is moved to A, and called B', as shown in Fig. 7. Once

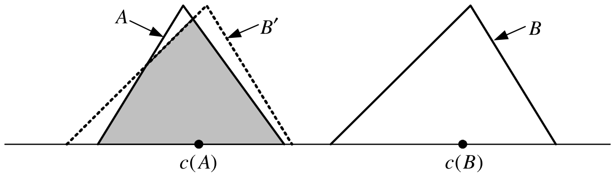

Fig. 7. An example of the VSM for T1 FSs. $c(A)$ and $c(B)$ are the centroids of A and B, respectively. B' is obtained by moving B so that $c(B)$ coincides with $c(A)$. Note that the shaded region can also be obtained by moving $c(A)$ to $c(B)$.

the two T1 FSs are "aligned", $s_1(A,B)$ is computed by replacing the IT2 FSs $\widetilde{A}$ and $\widetilde{B}$ in $s_1(\widetilde{A},\widetilde{B})$, given in (24) by T1 FSs A and B, i.e. by substituting $\overline{\mu}_{\tilde{A}} = \underline{\mu}_{\tilde{A}} = \mu_A$ and $\overline{\mu}_{\tilde{B}'} = \underline{\mu}_{\tilde{B}'} = \mu_{B'}$ into (24), so that

$$s_1(A,B) = \frac{\mathrm{card}(A \cap B')}{\mathrm{card}(A \cup B')} = \frac{\int_X \min(\mu_A(x), \mu_{B'}(x))\mathrm{d}x}{\int_X \max(\mu_A(x), \mu_{B'}(x))\mathrm{d}x}. \tag{29}$$

Note (29) is Jaccard's *unparameterized ratio model of similarity*[8] [17].

From Theorem 1, observe that $s_1(A,B)$ has the following properties: (1) $0 \leqslant s_1(A,B) \leqslant 1$; (2) $s_1(A,B) = 1 \iff A = B'$; and, (3) $s_1(A,B) = s_1(B,A)$.

5.2.2. Definition of $s_2(A,B)$

$s_2(A,B)$ measures the proximity of A and B. When IT2 FSs $\widetilde{A}$ and $\widetilde{B}$ become T1 FSs A and B, $s_2(\widetilde{A},\widetilde{B})$ in (25) becomes

$$s_2(A,B) = h(d(A,B)) \tag{30}$$

where

$$d(A,B) = |c(A) - c(B)| \tag{31}$$

is the Euclidean distance between the centroids of A and B (see Fig. 7). The definition of h is the same as the one given in Section 4.2.

Again, $s_2(A,B) \in [0,1]$. An example of $s_2(A,B)$ is

$$s_2(A,B) = \mathrm{e}^{-rd(A,B)}, \tag{32}$$

where r is a positive constant.

5.2.3. On converting $\mathbf{s}_v(A,B)$ to $s_s(A,B)$

The method proposed in Section 4.3 can also be used to convert $\mathbf{s}_v(A,B)$ to $s_s(A,B)$, i.e.

$$s_s(A,B) = s_1(A,B) \times s_2(A,B). \tag{33}$$

5.2.4. Comparison with Bonissone's linguistic approximation distance measure

Bonissone's [3,4] linguistic approximation distance measure was also proposed to identify the linguistic label which most closely resembles a given FS A; however, Bonissone modeled linguistic labels as T1 FSs, whereas we have modeled them as IT2 FSs.

The first step of Bonissone's method eliminates from further consideration those linguistic labels determined to be very far away from A. For a given T1 FS A, the distances between A and B_i, $d_1(A,B_i)$, are computed to identify M B_i that are close to A (according to some tolerance parameter). Bonissone [4] first computed four T1 FS features, *centroid*, *cardinality*, *fuzziness* and *skewness*, for A and B_i, and then defined $d_1(A,B_i)$ as the weighted Euclidean distance between the two four-dimensional points $[(p_A^1,p_A^2,p_A^3,p_A^4)^{\mathrm{T}}$ and $(p_{B_i}^1,p_{B_i}^2,p_{B_i}^3,p_{B_i}^4)^{\mathrm{T}}]$ represented by the values of the four features for each T1 FS, i.e.,

[8] It is called *coefficient of similarity* by Sneath in [39]. The term *index of communality* has also been used [8].

$$d_1(A, B_i) = \left[\sum_{j=1}^{4} w_j^2 (p_A^j - p_{B_i}^j)^2\right]^{1/2}. \tag{34}$$

The weights[9] w_j $(j = 1, 2, 3, 4)$ have to be pre-specified.

After pre-screening linguistic labels far away from A, Bonissone's second step uses the *modified Bhattacharya distance* [18] to discriminate between the M linguistic labels close to A, i.e.,

$$d_2(A, B_k) = \left[1 - \int_X \left(\frac{\mu_A(x)\mu_{B_k}(x)}{\text{card}(A)\cdot\text{card}(B_k)}\right)^{1/2} \mathrm{d}x\right]^{1/2} \quad k = 1, \ldots, M \tag{35}$$

The linguistic label corresponding to the smallest $d_2(A, B_k)$ is considered most similar to A.

Both $\mathbf{s}_v(A, B)$ and Bonissone's method consider the shapes and proximity of A and B. The main differences between them are:

(1) $\mathbf{s}_v(A, B)$ is a one-step method, whereas Bonissone's method is a two-step method.
(2) $\mathbf{s}_v(A, B)$ considers two features of A and B (shape and proximity). In Bonissone's first step, four features (centroid, cardinality, fuzziness and skewness) are considered, and in his second step, only one feature is considered (the modified Bhattacharya distance).
(3) $\mathbf{s}_v(A, B)$ measures the similarity between A and B, i.e. a larger $\mathbf{s}_v(A, B)$ means A and B are more similar. On the other hand, Bonissone's method measures the distance (or difference) between A and B, i.e. a larger $d_2(A, B)$ means A and B are less similar.

5.2.5. Comparison with Wenstøp's linguistic approximation method

Wenstøp [42], who considered the same problem as Bonissone, states: "*a linguistic approximation routine is a function from the set of fuzzy subsets to a set of linguistic values.*" Wenstøp used two parameters of a T1 FS, its *imprecision* (cardinality) and its *location* (centroid). The imprecision (p^1) was defined as the sum of membership values, whereas the location (p^2) was defined as the center of gravity. He then computed

$$d_W(A, B_i) = [(p_A^1 - p_{B_i}^1)^2 + (p_A^2 - p_{B_i}^2)^2]^{1/2} \quad i = 1, \ldots, N \tag{36}$$

and chose B_i with the smallest $d_W(A, B_i)$ as the one most similar to A. Observe that Wenstøp's method is a simplified version of Bonissone's first step, and his method is quite similar to the VSM method in that both of them use the centroid and cardinality. The differences are:

(1) The VSM computes the similarity between two T1 FSs, whereas Wenstøp's method computes the difference between two T1 FSs.
(2) The VSM first aligns A and B and then computes the cardinalities of $A \cap B'$ and $A \cup B'$, whereas Wenstøp's method computes cardinalities of A and B directly.
(3) The VSM can be used for T1 FSs of any shapes, whereas, as shown in [42], the two parameters in Wenstøp's method are insufficient criteria for satisfactory linguistic approximation. As a further refinement, he includes other characteristics of FSs, e.g. non-normality, multi-modality, fuzziness and dilation [42].

5.2.6. Comparison with Tsiporkova and Zimmermann's similarity measure

Tsiporkova and Zimmermann [40] proposed a similarity measure "*resulting from the aggregation of the compatibility and the equality of FSs*"

$$s_{TZ}(A, B) = \text{Agg}(\text{Com}(A, B), \text{Eql}(A, B)) \tag{37}$$

[9] We show w_j^2 in (34) rather than w_j, because this is the way the equation is stated in [4].

where $\mathrm{Com}(A,B)$ is a compatibility measure defined as

$$\mathrm{Com}(A,B) = \frac{\sup_{x\in X}(\mu_A(x)\bigstar\mu_B(x))}{\sup_{x\in X}(\mu_A(x)\vee\mu_B(x))}, \tag{38}$$

$\mathrm{Eql}(A,B)$ is an equality measure defined as

$$\mathrm{Eql}(A,B) = \mathrm{Inc}(A,B)\bigstar\mathrm{Inc}(B,A) \tag{39}$$

$$\mathrm{Inc}(A,B) = \inf_{x\in X} I(\mu_A(x),\mu_B(x)) \tag{40}$$

$$I(a,b) = \sup\{c|c\in[0,1] \text{ and } a\bigstar c\leqslant b\} \tag{41}$$

and $\mathrm{Agg}(a,b)$ is an aggregation operator defined as

$$\mathrm{Agg}(a,b) = a\bigstar(b\vee\lambda) \tag{42}$$

where λ is an adjustable parameter.

Tsiporkova and Zimmermann [40] did not specify an application for $s_{TZ}(A,B)$. It is of interest in this paper because $s_{TZ}(A,B)$ is quite similar to the VSM in that it also consists of two elements, and a crisp similarity is obtained by aggregating the two elements; however, the aggregation operator in $s_{TZ}(A,B)$ is more difficult to understand than the product operator used in the VSM. Additionally, special attention should be paid to the choice of λ in the aggregation operator. As pointed out by Tsiporkova and Zimmermann [40], "$\cdots$ *the choice of* λ *should be application-oriented. However, the use of a constant parameter for a compensation between the compatibility and the equality of FSs, does not seem to guarantee that the similarity measure will always be very sensitive to the different degrees of similarity or dissimilarity.*" For example, when $\mathrm{Com}(A,B)=1$ and $\mathrm{Eql}(A,B)=0$, the similarity is always λ (examples are shown in the next subsection). They continue to point out that "$\cdots$ *though such values for the compatibility and the equality can be obtained for many different pairs of FSs and it is rather strange to consider them similar to the same degree.*" Consequently, they suggested to use a dynamic λ, e.g., $\lambda = \mathrm{card}(A\cap B)/\mathrm{card}(A\cup B)$. Unfortunately, generally we lose the transitivity property of the similarity measure if dynamic λ is used [40].

5.2.7. Examples

For T1 FSs shown in Fig. 8, the results of Bonissone's linguistic approximation distance measure, Wenstøp's linguistic approximation measure, Tsiporkova and Zimmermann's similarity measure and the VSM are shown in Table 4. The domain of x was discretized into 201 equally-spaced points in all three methods. Note that all B_k $(k=1,\ldots,4)$ are assumed to survive Bonissone's first step, hence (35) was used to compute Bonissone's distance measure. Observe that Tsiporkova and Zimmermann's similarity measure with $\lambda=0.5$ indicates $s(A,B_1)=s(A,B_2)=\lambda=0.5$, which is counter-intuitive. This is an example of the problem pointed out at the end of Section 5.2.6. All other methods indicate B_2 is more similar to A than B_1 is, which seems reasonable.

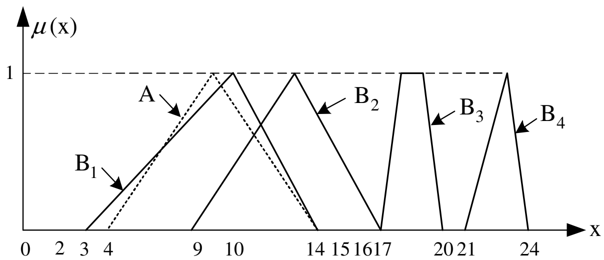

Fig. 8. T1 FSs used in the comparative study.

Table 4
Comparisons of distance/similarity measures for T1 FSs A and B_k ($k = 1, \ldots, 4$) shown in Fig. 8

Measure	$k=1$	$k=2$	$k=3$	$k=4$
$d_2(A, B_k)$	0.1086	0.7183	1	1
$d_W(A, B_k)$	9.0889	15.8994	20.6216	37.5736
$s_{TZ}(A, B_k)$ ($\lambda = 0.5$)	0.5	0.5	0	0
$s_{TZ}(A, B_k)$ $\left(\lambda_k = \frac{\text{card}(A\cap B_k)}{\text{card}(A\cup B_k)}\right)$ [a]	0.8394	0.1825	0	0
$s_s(A, B_k)$	0.8394	0.1615	0.0201	0.0013

Note that $d_2(A, B_k)$ and $d_W(A, B_k)$ are distance measures.
[a] In this case, $\lambda_1 = 0.8394$, $\lambda_2 = 0.1825$ and $\lambda_3 = \lambda_4 = 0$. Observe that $s_{TZ}(A, B_k) = \lambda_k$.

6. Conclusions

In this paper, our goal has been to solve the linguistic approximation problem, i.e., to find the $\widetilde{B}_i$ which most closely resembles $\widetilde{A}$. After reviewing four existing similarity measures for IT2 FSs and pointing out their short-comings, a vector similarity measure for IT2 FSs was proposed. Because a T1 FS is a special cases of an IT2 FS (when all uncertainty disappears), the proposed VSM can also be used for T1 FSs.

The VSM is the first IT2 FS similarity measure that has a vector form. It is easy to understand, and its two components enable us to consider the similarity between shapes and proximity separately and explicitly. One reviewer mentioned the problem of looking for a particular shape in a figure. In this context, proximity and rotation are not important, and only the shape should be considered. The VSM cannot handle this case because it is a similarity measure of FSs, and it may not be useful as a similarity measure of points, vectors, figures, functions, etc. Our comparative study showed that the VSM gives reasonable similarity measures in linguistic approximation and does not have the short-comings of the four existing similarity measures. It has already been used in [43] as the decoder in the Per–C (Fig. 1).

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments which helped us to improve this paper.

Appendix A. Background on interval type-2 fuzzy sets

A.1. Interval type-2 fuzzy sets (IT2 FS)

An IT2 FS, $\widetilde{A}$, is to-date the most widely used kind of T2 FS, and is the only kind of T2 FS that is considered in this paper. It is described as[10]

$$\widetilde{A} = \int_{x\in X} \int_{u\in J_x} 1/(x,u) = \int_{x\in X} \left[\int_{u\in J_x} 1/u\right] \Big/ x \tag{A.1}$$

where x is the *primary variable*, J_x, an interval in $[0,1]$, is the *primary membership* of x, u is the *secondary variable*, and $\int_{u\in J_x} 1/u$ is the *secondary membership function* (MF) at x. Note that (A.1) means: $\widetilde{A} : X \to \{[a,b] : 0 \leqslant a \leqslant b \leqslant 1\}$. Uncertainty about $\widetilde{A}$ is conveyed by the union of all of the primary memberships, called the *footprint of uncertainty* of $\widetilde{A}$ [FOU($\widetilde{A}$)], i.e.

$$\text{FOU}(\widetilde{A}) = \bigcup_{x\in X} J_x = \bigcup_{x\in X} [\underline{\mu}_{\widetilde{A}}(x), \overline{\mu}_{\widetilde{A}}(x)] \tag{A.2}$$

[10] This background material is taken from [33]. See also [25].

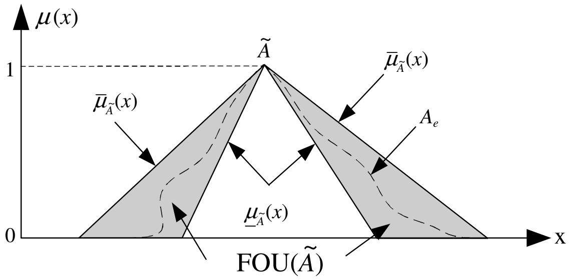

Fig. 9. An interval type-2 fuzzy set. A_e is an embedded type-1 fuzzy set.

An IT2 FS is shown in Fig. 9. The FOU is shown as the shaded region. It is bounded by an *upper MF* (UMF) $\overline{\mu}_{\widetilde{A}}(x)$ and a *lower MF* (LMF) $\underline{\mu}_{\widetilde{A}}(x)$, both of which are T1 FSs; consequently, the membership grade of each element of an IT2 FS is an interval $[\underline{\mu}_{\widetilde{A}}(x), \overline{\mu}_{\widetilde{A}}(x)]$.

Note that an IT2 FS can also be represented as

$$\widetilde{A} = 1/\mathrm{FOU}(\widetilde{A}) \tag{A.3}$$

with the understanding that this means putting a secondary grade of 1 at all points of $\mathrm{FOU}(\widetilde{A})$.

For discrete universes of discourse X and U, an *embedded T1 FS* A_e has N elements, one each from $J_{x_1}, J_{x_2}, \ldots, J_{x_N}$, namely $u_1, u_2, \ldots, u_N$, i.e.

$$A_e = \sum_{i=1}^{N} u_i/x_i \quad u_i \in J_{x_i} \subseteq U = [0,1] \tag{A.4}$$

Examples of A_e are $\overline{\mu}_{\widetilde{A}}(x)$ and $\underline{\mu}_{\widetilde{A}}(x)$; see, also Fig. 9. Note that if each u_i is discretized into M_i levels, there will be a total of n_A A_e, where

$$n_A = \prod_{i=1}^{N} M_i \tag{A.5}$$

Mendel and John [29] have presented a Representation Theorem for a general T2 FS, which when specialized to an IT2 FS can be expressed as:

Representation Theorem for an IT2 FS: Assume that primary variable x of an IT2 FS $\widetilde{A}$ is sampled at N values, $x_1, x_2, \ldots, x_N$, and at each of these values its primary memberships u_i are sampled at M_i values, $u_{i1}, u_{i2}, \ldots, u_{iM_i}$. Let A_e^j denote the jth embedded T1 FS for $\widetilde{A}$. Then $\widetilde{A}$ is represented by (A.3), in which[11]

$$\mathrm{FOU}(\widetilde{A}) = \bigcup_{j=1}^{n_A} A_e^j = \{\underline{\mu}_{\widetilde{A}}(x), \ldots, \overline{\mu}_{\widetilde{A}}(x)\} \equiv [\underline{\mu}_{\widetilde{A}}(x), \overline{\mu}_{\widetilde{A}}(x)]. \tag{A.6}$$

This representation of an IT2 FS, in terms of simple T1 FSs, the embedded T1 FSs, is very useful for deriving theoretical results; however, it is not recommended for computational purposes, because it would require the enumeration of the n_A embedded T1 FSs and n_A [given in (A.5)] can be astronomical.

A.2. Centroid of an IT2 FS

The centroid of an IT2 FS has been well-defined by Karnik and Mendel [19]. Let A_e be an embedded T1 FS of an IT2 FS $\widetilde{A}$. The centroid of $\widetilde{A}$ is defined as the union of the centroids of all A_e, i.e.,

$$C_{\widetilde{A}} \equiv \bigcup_{\forall A_e} c(A_e) = \bigcup_{\forall A_e} \frac{\int_X x \cdot \mu_{A_e}(x)\mathrm{d}x}{\int_X \mu_{A_e}(x)\mathrm{d}x} = [c_l(\widetilde{A}), c_r(\widetilde{A})] \tag{A.7}$$

[11] Although there are a finite number of embedded T1 FSs, it is customary to represent $\mathrm{FOU}(\widetilde{A})$ as an interval set $[\underline{\mu}_{\widetilde{A}}(x), \overline{\mu}_{\widetilde{A}}(x)]$ at each x. Doing this is equivalent to discretizing with infinitesimally many small values and letting the discretizations approach zero.

where $c(A_e)$ is the centroid of A_e, and $c_l(\widetilde{A})$ and $c_r(\widetilde{A})$ are the minimum and maximum centroids of all A_e, respectively. $c_l(\widetilde{A})$ and $c_r(\widetilde{A})$ can be computed by using the Karnik–Mendel (KM) algorithms [25].

A.3. Cardinality of an IT2 FS

Definitions of the cardinality of T1 FSs have been proposed by several authors, e.g. De Luca and Termini [9], Kaufmann [21], Gottwald [14], Zadeh [47], Blanchard [2], Klement [23], Wygralak [44], etc. Basically there are two kinds of proposals [11]: (1) those which assume that the cardinality of a T1 FS can be a crisp number; and, (2) those which claim that it should be a fuzzy number. De Luca and Termini's definition [9] is the most frequently used definition of cardinality for T1 FSs:

$$\mathrm{card}(A) = \int_X \mu_A(x)\mathrm{d}x. \tag{A.8}$$

It is adopted in this paper.

The cardinality of an IT2 FS $\widetilde{A}$ is defined as the union of all cardinalities of its embedded T1 FSs A_e, i.e.,

$$\mathrm{card}(\widetilde{A}) \equiv \bigcup_{\forall A_e} \mathrm{card}(A_e) = \bigcup_{\forall A_e} \int_X \mu_{A_e}(x)\mathrm{d}x = \left[\min_{\forall A_e} \int_X \mu_{A_e}(x)\mathrm{d}x, \max_{\forall A_e} \int_X \mu_{A_e}(x)\mathrm{d}x\right]. \tag{A.9}$$

(A.9) can be easily computed by:

Theorem A.1: $\mathrm{card}(\widetilde{A})$ in (A.9) can be re-expressed as

$$\mathrm{card}(\widetilde{A}) = [\mathrm{card}(\underline{\mu}_{\tilde{A}}(x)), \mathrm{card}(\overline{\mu}_{\tilde{A}}(x))]. \tag{A.10}$$

Proof

$$\min_{\forall A_e} \int_X \mu_{A_e}(x)\mathrm{d}x = \int_X [\min_{\forall A_e} \mu_{A_e}(x)]\mathrm{d}x = \int_X \underline{\mu}_{\tilde{A}}(x)\mathrm{d}x = \mathrm{card}(\underline{\mu}_{\tilde{A}}(x)) \tag{A.11}$$

$$\max_{\forall A_e} \int_X \mu_{A_e}(x)\mathrm{d}x = \int_X [\max_{\forall A_e} \mu_{A_e}(x)]\mathrm{d}x = \int_X \overline{\mu}_{\tilde{A}}(x)\mathrm{d}x = \mathrm{card}(\overline{\mu}_{\tilde{A}}(x)). \tag{A.12}$$

(A.10) is obtained by substituting (A.11) and (A.12) into (A.9). □

Additionally, we define the *average cardinality* of $\widetilde{A}$ as the average of its minimum and maximum cardinalities, i.e.,

$$\mathrm{AC}(\widetilde{A}) = \frac{\mathrm{card}(\underline{\mu}_{\tilde{A}}(x)) + \mathrm{card}(\overline{\mu}_{\tilde{A}}(x))}{2}. \tag{A.13}$$

Appendix B. Proof of theorems

B.1. Proof of theorem 1

B.1.1. Proof of (a)

Because

$$0 \leqslant \min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}'}(x)) \leqslant \max(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}'}(x)) \tag{B.1}$$

$$0 \leqslant \min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}'}(x)) \leqslant \max(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}'}(x)), \tag{B.2}$$

it follows that

$$0 \leqslant \int_X \min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \leqslant \int_X \max(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \tag{B.3}$$

$$0 \leqslant \int_X \min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \leqslant \int_X \max(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \tag{B.4}$$

Consequently,

$$s_1(\widetilde{A},\widetilde{B}) = \frac{\int_X \min(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x + \int_X \min(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x}{\int_X \max(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x + \int_X \max(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x} \in [0,1]. \tag{B.5}$$

B.1.2. Proof of (b)

$\widetilde{A} = \widetilde{B}'$ means $\underline{\mu}_{\tilde{A}}(x) = \underline{\mu}_{\tilde{B}'}(x)$ and $\overline{\mu}_{\tilde{A}}(x) = \overline{\mu}_{\tilde{B}'}(x)$ for $\forall x \in X$. Substituting these two equations into (24),

$$s_1(\widetilde{A},\widetilde{B}) = \frac{\int_X \overline{\mu}_{\tilde{A}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\tilde{A}}(x)\mathrm{d}x}{\int_X \overline{\mu}_{\tilde{A}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\tilde{A}}(x)\mathrm{d}x} = 1, \tag{B.6}$$

which proves the necessity of Theorem 1b.

To prove the sufficiency of the result, observe, from (24), that $s_1(\widetilde{A},\widetilde{B}) = 1$ means

$$\begin{aligned}&\int_X \min(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x + \int_X \min(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x\\ &\quad = \int_X \max(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x + \int_X \max(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x.\end{aligned} \tag{B.7}$$

For IT2 FSs $\widetilde{A}$ and $\widetilde{B}$, $\int_X \min(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \neq \int_X \max(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x$ and $\int_X \min(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \neq \int_X \max(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x$. So, (B.7) holds only when

$$\int_X \min(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x = \int_X \max(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \tag{B.8}$$

$$\int_X \min(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x = \int_X \max(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x))\mathrm{d}x \tag{B.9}$$

(B.8) holds if and only if

$$\overline{\mu}_{\tilde{A}}(x) = \overline{\mu}_{\tilde{B}'}(x) \quad \forall x \in X. \tag{B.10}$$

(B.9) holds if and only if

$$\underline{\mu}_{\tilde{A}}(x) = \underline{\mu}_{\tilde{B}'}(x) \quad \forall x \in X. \tag{B.11}$$

(B.10) and (B.11) together mean $\widetilde{A} = \widetilde{B}'$.

B.1.3. Proof of (c)

$s_1(\widetilde{A},\widetilde{B}) = s_1(\widetilde{B},\widetilde{A})$ is obvious because the min and max operators in (24) do not concern the order of $\overline{\mu}_{\tilde{A}}(x)$ and $\overline{\mu}_{\tilde{B}'}(x)$, i.e. $\min(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x)) = \min(\overline{\mu}_{\tilde{B}'}(x),\overline{\mu}_{\tilde{A}}(x))$ and $\max(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}'}(x)) = \max(\overline{\mu}_{\tilde{B}'}(x),\overline{\mu}_{\tilde{A}}(x))$, and, they do not concern the order of $\underline{\mu}_{\tilde{A}}(x)$ and $\underline{\mu}_{\tilde{B}'}(x)$ either, i.e. $\min(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x)) = \min(\underline{\mu}_{\tilde{B}'}(x),\underline{\mu}_{\tilde{A}}(x))$ and $\max(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}'}(x)) = \max(\underline{\mu}_{\tilde{B}'}(x),\underline{\mu}_{\tilde{A}}(x))$. □

B.2. Proof of theorem 3

B.2.1. Proof of (a)

Sufficiency: $\widetilde{A} = \widetilde{B}$ means $s_1(\widetilde{A},\widetilde{B}) = 1$ and $s_2(\widetilde{A},\widetilde{B}) = 1$; hence, $s_s(\widetilde{A},\widetilde{B}) = 1$.

Necessity: $s_s(\widetilde{A},\widetilde{B}) = 1$ if and only if $s_1(\widetilde{A},\widetilde{B}) = 1$ and $s_2(\widetilde{A},\widetilde{B}) = 1$. $s_1(\widetilde{A},\widetilde{B}) = 1$ means the shapes of $\widetilde{A}$ and $\widetilde{B}$ are the same, and $s_2(\widetilde{A},\widetilde{B}) = 1$ means the distance between $\widetilde{A}$ and $\widetilde{B}$ is zero. Consequently, $\widetilde{A} = \widetilde{B}$. □

B.2.2. Proof of (b)

Because neither $s_1(\widetilde{A},\widetilde{B})$ nor $s_2(\widetilde{A},\widetilde{B})$ concern the order of $\widetilde{A}$ and $\widetilde{B}$, i.e. $s_1(\widetilde{A},\widetilde{B}) = s_1(\widetilde{B},\widetilde{A})$ and $s_2(\widetilde{A},\widetilde{B}) = s_2(\widetilde{B},\widetilde{A})$, it follows that $s_s(\widetilde{A},\widetilde{B}) = s_s(\widetilde{B},\widetilde{A})$. □

B.2.3. Proof of (c)

$\widetilde{A}$, $\widetilde{B}$ and $\widetilde{C}$ having the same shape means

$$s_1(\widetilde{A},\widetilde{B}) = s_1(\widetilde{A},\widetilde{C}) = s_1(\widetilde{B},\widetilde{C}) = 1. \tag{B.12}$$

$c(\widetilde{A}) < c(\widetilde{B}) < c(\widetilde{C})$ means $d(\widetilde{A},\widetilde{B}) < d(\widetilde{A},\widetilde{C})$ and $d(\widetilde{B},\widetilde{C}) < d(\widetilde{A},\widetilde{C})$; consequently,

$$s_2(\widetilde{A},\widetilde{B}) > s_2(\widetilde{A},\widetilde{C}). \tag{B.13}$$

and

$$s_2(\widetilde{B},\widetilde{C}) > s_2(\widetilde{A},\widetilde{C}). \tag{B.14}$$

Hence,

$$s_1(\widetilde{A},\widetilde{B}) \times s_2(\widetilde{A},\widetilde{B}) > s_1(\widetilde{A},\widetilde{C}) \times s_2(\widetilde{A},\widetilde{C}), \tag{B.15}$$

and

$$s_1(\widetilde{B},\widetilde{C}) \times s_2(\widetilde{B},\widetilde{C}) > s_1(\widetilde{A},\widetilde{C}) \times s_2(\widetilde{A},\widetilde{C}), \tag{B.16}$$

i.e., $s_s(\widetilde{A},\widetilde{B}) > s_s(\widetilde{A},\widetilde{C})$ and $s_s(\widetilde{B},\widetilde{C}) > s_s(\widetilde{A},\widetilde{C})$. □

B.2.4. Proof of (d)

$c(\widetilde{A}) = c(\widetilde{B}) = c(\widetilde{C})$ means

$$s_2(\widetilde{A},\widetilde{B}) = s_2(\widetilde{A},\widetilde{C}) = s_2(\widetilde{B},\widetilde{C}) = 1. \tag{B.17}$$

$\widetilde{A} < \widetilde{B}$ means

$$s_1(\widetilde{A},\widetilde{B}) = \frac{\int_X \min(\overline{\mu}_{\widetilde{A}}(x),\overline{\mu}_{\widetilde{B}}(x))\mathrm{d}x + \int_X \min(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}}(x))\mathrm{d}x}{\int_X \max(\overline{\mu}_{\widetilde{A}}(x),\overline{\mu}_{\widetilde{B}}(x))\mathrm{d}x + \int_X \max(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}}(x))\mathrm{d}x} = \frac{\int_X \overline{\mu}_{\widetilde{A}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\widetilde{A}}(x)\mathrm{d}x}{\int_X \overline{\mu}_{\widetilde{B}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\widetilde{B}}(x)\mathrm{d}x}; \tag{B.18}$$

$\widetilde{A} < \widetilde{C}$ means

$$s_1(\widetilde{A},\widetilde{C}) = \frac{\int_X \overline{\mu}_{\widetilde{A}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\widetilde{A}}(x)\mathrm{d}x}{\int_X \overline{\mu}_{\widetilde{C}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\widetilde{C}}(x)\mathrm{d}x}; \tag{B.19}$$

and, $\widetilde{B} < \widetilde{C}$ means

$$\int_X \overline{\mu}_{\widetilde{B}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\widetilde{B}}(x)\mathrm{d}x < \int_X \overline{\mu}_{\widetilde{C}}(x)\mathrm{d}x + \int_X \underline{\mu}_{\widetilde{C}}(x)\mathrm{d}x \tag{B.20}$$

Substituting (B.20) into (B.18) and (B.19), it follows that $s_1(\widetilde{A},\widetilde{B}) > s_1(\widetilde{A},\widetilde{C})$; consequently, $s_s(\widetilde{A},\widetilde{B}) > s_s(\widetilde{A},\widetilde{C})$. Similarly, we can prove $s_s(\widetilde{B},\widetilde{C}) > s_s(\widetilde{A},\widetilde{C})$. □

B.2.5. Proof of (e)

Observe that $s_1(\widetilde{A},\widetilde{B}) > 0$ and $s_2(\widetilde{A},\widetilde{B}) > 0$. Consequently, $s_s(\widetilde{A},\widetilde{B}) > 0$. □

References

[1] A. Bhattacharya, On a measure of divergence of two multinomial populations, Sankhya 7 (1946) 401–406.
[2] N. Blanchard, Cardinal and ordinal theories about fuzzy sets, in: M.M. Gupta, E. Sanchez (Eds.), Fuzzy Information and Decision Processes, North-Holland, Amsterdam, 1982, pp. 149–157.
[3] P.P. Bonissone, A pattern recognition approach to the problem of linguistic approximation, in: Proceedings of the IEEE International Conference On Cybernetics and Society, Denver, CO, October 1979, pp. 793–798.
[4] P.P. Bonissone, A fuzzy sets based linguistic approach: Theory and applications, in: Proceedings of the 12th Winter Simulation Conference, Orlando, FL, 1980, pp. 99–111.
[5] P.P. Bonissone, A. Varma, K.S. Aggour, F. Xue, Design of local fuzzy models using evolutionary algorithms, Journal of Computational Statistics and Data Analysis 51 (1) (2006) 398–416.
[6] H. Bustince, Indicator of inclusion grade for interval-valued fuzzy sets. application to approximate reasoning based on interval-valued fuzzy sets, International Journal of Approximate Reasoning 23 (3) (2000) 137–209.

[7] H. Bustince, M. Pagola, E. Barrenechea, Construction of fuzzy indices from fuzzy DI-subsethood measures: application to the global comparison of images, Information Sciences 177 (2007) 906–929.
[8] V.V. Cross, T.A. Sudkamp, Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications, Physica-Verlag, Heidelberg, NY, 2002.
[9] A. De Luca, S. Termini, A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, Information and Computation 20 (1972) 301–312.
[10] D. Dubois, H. Prade, A unifying view of comparison indices in a fuzzy set-theoretic framework, in: R. Yager (Ed.), Fuzzy Set and Possibility Theory Recent Developments, Pergamon Press, NY, 1982, pp. 3–13.
[11] D. Dubois, H. Prade, Fuzzy cardinality and the modeling of imprecise quantification, Fuzzy Sets and Systems 16 (1985) 199–230.
[12] J. Fan, W. Xie, Some notes on similarity measure and proximity measure, Fuzzy Sets and Systems 101 (1999) 403–412.
[13] M.B. Gorzalczany, A method of inference in approximate reasoning based on interval-valued fuzzy sets, Fuzzy Sets and Systems 21 (1987) 1–17.
[14] S. Gottwald, A note on fuzzy cardinals, Kybernetika 16 (1980) 156–158.
[15] K. Hirota, W. Pedrycz, Handling fuzziness and randomness in process of matching fuzzy data, in: Proceedings of the Third IFSA Congress, 1989, pp. 97–100.
[16] K. Hirota, W. Pedrycz, Matching fuzzy quantities, IEEE Transactions on Systems, Man and Cybernetics 21 (6) (1991) 908–914.
[17] P. Jaccard, Nouvelles recherches sur la distribution florale, Bulletin de la Societe de Vaud des Sciences Naturelles 44 (1908) 223.
[18] T. Kailath, The divergence and bhattacharyaa distance measure in signal detection, IEEE Transactions on Communication Technology 15 (1967) 609–637.
[19] N.N. Karnik, J.M. Mendel, Centroid of a type-2 fuzzy set, Information Sciences 132 (2001) 195–220.
[20] A. Kaufmann, Introduction to the Theory of Fuzzy Sets, Academic Press, NY, 1975.
[21] A. Kaufmann, Introduction a la theorie des sous-ensembles flous, Complement et Nouvelles Applications, vol. 4, Masson, Paris, 1977.
[22] F. Klawonn, Should fuzzy equality and similarity satisfy transitivity? comments on the paper by m. de cock and e. kerre, Fuzzy Sets and Systems 133 (2003) 175–180.
[23] E.P. Klement, On the cardinality of fuzzy sets, in: Proceedings of the Sixth European Meeting on Cybernetics and Systems Research, Vienna, 1982, pp. 701–704.
[24] C. Mencar, G. Castellano, A.M. Fanelli, Distinguishability quantification of fuzzy sets, Information Sciences 177 (2007) 130–149.
[25] J.M. Mendel, Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice-Hall, Upper Saddle River, NJ, 2001.
[26] J.M. Mendel, Computing with words, when words can mean different things to different people, in: Proceedings of the Third International ICSC Symposium on Fuzzy Logic and Applications, Rochester, NY, June 1999, pp. 158–164.
[27] J.M. Mendel, An architecture for making judgement using computing with words, International Journal of Applied Mathematics and Computer Science 12 (3) (2002) 325–335.
[28] J.M. Mendel, Computing with words and its relationships with fuzzistics, Information Sciences 177 (2007) 988–1006.
[29] J.M. Mendel, R.I. John, Type-2 fuzzy sets made simple, IEEE Transactions on Fuzzy Systems 10 (2) (2002) 117–127. April.
[30] J.M. Mendel, H. Wu, Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 1, forward problems, IEEE Transactions on Fuzzy Systems 14 (6) (2006) 781–792.
[31] J.M. Mendel, H. Wu, Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 2, inverse problems, IEEE Transactions on Fuzzy Systems 15 (2) (2007) 301–308.
[32] J.M. Mendel, H. Wu, Centroid uncertainty bounds for interval type-2 fuzzy sets: forward and inverse problems, in: Proceedings of the FUZZ-IEEE, vol. 2, Budapest, Hungary, July 2004, pp. 947–952.
[33] J.M. Mendel, H. Hagras, R.I. John, Standard background material about interval type-2 fuzzy logic systems that can be used by all authors, http://ieee-cis.org/_files/standards.t2.win.pdf.
[34] H.B. Mitchell, Pattern recognition using type-II fuzzy sets, Information Sciences 170 (2005) 409–418.
[35] M. Nikravesh, Soft computing for reservoir characterization and management, in: Proceedings of the IEEE International Conference on Granular Computing, vol. 2, Beijing, China, July 2005, pp. 593–598.
[36] S.K. Pal, M.M. Majumder, Fuzzy sets and decision-making approaches in vowel and speaker recognition, IEEE Transactions on Systems, Man and Cybernetics 7 (1977) 625–629.
[37] A.L. Ralescu, D.A. Ralescu, Probability and fuzzyiness, Information Sciences 34 (1984) 85–92.
[38] M. Setnes, R. Babuska, U. Kaymak, H.R. Van, N. Lemke, Similarity measures in fuzzy rule base simplification, IEEE Transactions on Systems, Man and Cybernetics, Part B 28 (3) (1998) 376–386.
[39] P.H.A. Sneath, R.R. Sokal, Numerical Taxonomy, W.H. Freeman and Company, San Francisco, CA, 1973.
[40] E. Tsiporkova, H.-J. Zimmermann, Aggregation of compatibility and equality: a new class of similarity measures for fuzzy sets, in: Proceedings of the Seventh International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris, 1998, pp. 1769–1776.
[41] A. Tversky, Features of similarity, Psychology Review 84 (1977) 327–352.
[42] F. Wenstøp, Quantitative analysis with linguistic values, Fuzzy Sets and Systems 4 (1980) 99–115.
[43] D. Wu, J.M. Mendel, Aggregation using the linguistic weighted average and interval type-2 fuzzy sets, IEEE Transactions on Fuzzy Systems, 2007, in press.
[44] M. Wygralak, A new approach to the fuzzy cardinality of finite fuzzy sets, Busefal 15 (1983) 72–75.
[45] R.R. Yager, A framework for multi-source data fusion, Information Sciences 163 (2004) 175–200.
[46] L.A. Zadeh, Similarity relations and fuzzy orderings, Information Sciences 3 (1971) 177–200.

[47] L.A. Zadeh, Possibility theory and soft data analysis, in: L. Cobb, R.M. Thrall (Eds.), Mathematical Frontiers of the Social and Policy Sciences, Westview Press, Boulder, CO, 1981, pp. 69–129.
[48] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Transactions on Fuzzy Systems 4 (1996) 103–111.
[49] L.A. Zadeh, From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions, IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications 4 (1999) 105–119.
[50] L.A. Zadeh, Toward a generalized theory of uncertainty (GTU) – an outline, Information Sciences 172 (2005) 1–40.
[51] W. Zeng, H. Li, Relationship between similarity measure and entropy of interval valued fuzzy sets, Fuzzy Sets and Systems 157 (2006) 1477–1484.
[52] R. Zwick, E. Carlstein, D. Budescu, Measures of similarity among fuzzy concepts: a comparative analysis, International Journal of Approximate Reasoning 1 (1987) 221–242.