USC-SIPI REPORT #398

Interval Type-2 Fuzzy Set Subsethood Measures as a Decoder for Perceptual Computing

by

Dongrui Wu

December 2009

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Viterbi School of Engineering
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

Interval Type-2 Fuzzy Set Subsethood Measures as A Decoder for Perceptual Computing

Dongrui Wu and Jerry M. Mendel

Signal and Image Processing Institute
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089-2564
Tel: (213)740-4445, Fax: (213)740-4651
Email: dongruiw@usc.edu
mendel@sipi.usc.edu

January 30, 2010

Abstract

In some applications of computing with words, it is necessary to map an interval type-2 fuzzy set (IT2 FS) into one of several classes, which are also represented by IT2 FSs. This classifier can be implemented by a subsethood measure. Five existing subsethood measures for IT2 FSs are considered in this paper. Comparative studies show that Vlachos and Sergiadis's IT2 FS subsethood measure gives the most reasonable outputs as a decoder in computing with words when the desired output is a class. The results in this paper will be useful in constructing a third kind of decoder (i.e., in addition to similarity measures and ranking methods) for perceptual computing.

Keywords: Computing with words, interval type-2 fuzzy sets, perceptual computing, retranslation, similarity measure, subsethood measure

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Zadeh coined the phrase "*computing with words*" (CWW) [66, 67], which is [67] *"a methodology in which the objects of computation are words and propositions drawn from a natural language."* Words in the CWW paradigm may be modeled by type-1 fuzzy sets (T1 FSs) [22, 64] or their extension, interval type-2 (IT2) FSs [22, 26–28, 65]. So, an inevitable question is: Which FS model should be used in CWW?

There are at least two types of uncertainties associated with a word [24, 42]: *intra-personal uncertainty* and *inter-personal uncertainty*. Intra-personal uncertainty describes [24] "*the uncertainty a person has about the word.*" It is also explicitly pointed out by Wallsten and Budescu [42] as *"except in very special cases, all representations are vague to some degree in the minds of the originators and in the minds of the receivers,"* and they suggest to model it by a T1 FS. Inter-personal uncertainty describes [24] "*the uncertainty that a group of people have about the word.*" It is pointed out by Mendel [22] as *"words mean different things to different people"* and Wallsten and Budescu [42] as *"different individuals use diverse expressions to describe identical situations and understand the same phrases differently when hearing or reading them."* Because an IT2 FS has a footprint of uncertainty (FOU, e.g., Fig. 4.1(a)) which can be viewed as a group of T1 FSs, it can model both types of uncertainty [24]; hence, we suggest IT2 FSs be used in CWW [21, 22, 24]. Additionally, Mendel [25] has explained why it is scientifically incorrect to model a word using a T1 FS, i.e., (1) A T1 FS for a word is well-defined by its membership function (MF) that is totally certain once all of its parameters are specified; (2) words mean different things to different people, and so are uncertain; and, therefore, (3) it is a contradiction to say that something certain can model something that is uncertain.

CWW using T1 FSs has been studied by many researchers, e.g. [7, 9, 13, 18, 34, 37, 40, 43–45, 56, 62, 67, 70]; however, because of the above arguments, in this paper IT2 FSs are used to model words.

A specific architecture, proposed in [23] for making subjective judgments by CWW, is shown in Fig. 1.1. It is called a *Perceptual Computer*—Per-C for short. In Fig. 1.1, the *encoder*[1] transforms linguistic perceptions into IT2 FSs that activate a *CWW engine*[2]. The CWW engine performs oper-

[1]Zadeh calls this *constraint explicitation* in [66, 67]. In [68, 69] and some of his recent talks, he calls this *precisiation*. Yager [56] and Martin and Klir [19] call this *translation*.

[2]Zadeh calls this *constraint propagation* in [66, 67]. Yager [56] calls this *inference/granular computing*. Martin and Klir [19] call this *approximate reasoning*.

ations on the IT2 FSs. The *decoder*[3] maps the output of the CWW engine into a recommendation, which can be a word, rank, or class.

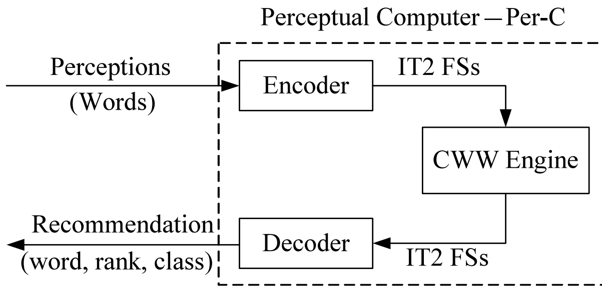

Fig. 1.1: Conceptual structure of the Perceptual Computer.

To operate the Per-C, one needs to solve the following problems:

1. *How to transform words into IT2 FSs, i.e., the encoding problem.* This can be done with Liu and Mendel's *Interval Approach* (IA) [15]. First, for each word in an application-dependent encoding vocabulary, a group of subjects are asked the following question:

   On a scale of 0-10, what are the end-points of an interval that you associate with the word ____?

   After some pre-processing, during which some intervals (e.g., outliers) are eliminated, each of the remaining intervals is classified as either an interior, left-shoulder or right-shoulder IT2 FS. Then, each of the word's data intervals is individually mapped into its respective T1 interior, left-shoulder or right-shoulder MF, after which the union of all of these T1 MFs is taken. The result is an FOU for an IT2 FS model of the word. The words and their FOUs constitute a *codebook*.

2. *How to construct the CWW engine, which maps IT2 FSs into IT2 FSs.* There are different kinds of CWW engines, e.g.,

   (a) The *linguistic weighted average* [46–48, 50], which is defined as

   $$\tilde{Y} = \frac{\sum_{i=1}^{N} \tilde{X}_i \tilde{W}_i}{\sum_{i=1}^{N} \tilde{W}_i} \tag{1.1}$$

   where $\tilde{X}_i$, the sub-criteria (e.g., data, features, decisions, recommendations, judgments, scores, etc), and $\tilde{W}_i$, the weights, are usually words modeled by IT2 FSs; however, they can also be special cases of IT2 FSs, e.g., numbers, intervals, or T1 FSs. It is shown [46–48, 50] that the UMF of $\tilde{Y}$ is a fuzzy weighted average [14] of the UMFs of $\tilde{X}_i$ and $\tilde{W}_i$, and the LMF of $\tilde{Y}$ is a fuzzy weighted average of the LMFs of $\tilde{X}_i$ and $\tilde{W}_i$.

[3]Zadeh calls this *linguistic approximation* or *constraint retranslation* in [66, 67]. Yager [56] and Martin and Klir [19] call this *retranslation*.

(b) *Perceptual reasoning (PR)* [29, 46, 53, 54], which considers the following problem:
Given a rulebase with K rules, each of the form:

$$R^k : \textit{If } x_1 \textit{ is } \tilde{F}_1^k \textit{ and } \dots \textit{ and } x_p \textit{ is } \tilde{F}_p^k, \textit{ Then } y \textit{ is } \tilde{G}^k \tag{1.2}$$

where $\tilde{F}_j^k$ and $\tilde{G}^k$ are words modeled by IT2 FSs, and a new input $\tilde{\mathbf{X}}' = (\tilde{X}_1, \dots, \tilde{X}_p)$, where $\tilde{X}_j$ $(j = 1, \dots, p)$ are also words modeled by IT2 FSs, then what is the output IT2 FS $\tilde{Y}_{PR}$?
In similarity-based PR [46, 53, 54] one computes

$$\tilde{Y}_{PR} = \frac{\sum_{k=1}^{K} f^k(\tilde{\mathbf{X}}')\tilde{G}^k}{\sum_{k=1}^{K} f^k(\tilde{\mathbf{X}}')} \tag{1.3}$$

where $f^k(\tilde{\mathbf{X}}')$ is the firing level of R^k, i.e.,

$$f^k(\tilde{\mathbf{X}}') = \prod_{j=1}^{p} sm_J(\tilde{X}_j, \tilde{F}_j^k) \tag{1.4}$$

in which sm_J is the Jaccard similarity for IT2 FSs [51], i.e.,

$$sm_J(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{N} \max(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N} \max(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))}. \tag{1.5}$$

Another approach that uses firing intervals instead of firing levels is described in [29].

3. *How to map the output of the CWW engine into a recommendation, i.e., the decoding problem.* This is the focus of this paper, and a method using subsethood is proposed.

The rest of this paper is organized as follows: Section 2 describes different decoding problems in detail. Section 3 introduces five existing subsethood measures of IT2 FSs and their properties. Section 4 compares the performance of the five subsethood measures, as well as the Jaccard similarity measure, as a decoder when the desired output is a class. Finally, Section 5 draws conclusions.

Chapter 2

The Decoding Problem

The decoding problem is challenging because there can be different forms of recommendations, and hence different kinds of decoders are needed. Three different forms of recommendations and their corresponding decoders are introduced next.

2.1 Rank

In some decision-making situations several alternatives are compared at the same time so that the best one can be chosen.

Example 1 *In [46, 52] a contractor has to decide which of three companies (A, B or C) is going to get the final mass production contract for a missile system. In the CWW engine, after aggregation across different criteria and different levels, the overall performance of each missile system is represented by an IT2 FS. The decoder needs to rank them to find the best system.* ■

The centroid-based ranking method for IT2 FSs [51] can be used when the recommendation is a rank.

When ranking alternatives, it may not be necessary to map each IT2 FS that is output by the CWW engine into a word for understanding; however, if this mapping is needed, then the following *word* decoder should also be used in addition to the *rank* decoder.

2.2 Word

This is the most typical and also most frequently studied case [19, 56, 66, 67], where the IT2 FS output by the CWW engine has to be mapped into a word in the *same* vocabulary as the one used by the CWW engine so that the IT2 FS can be understood.

Example 2 *In the journal publication judgment advisor [30, 46], the overall quality of a paper is evaluated using two major criteria (*Technical Merit *and* Presentation*), and each major criterion also has some sub-criteria (*Technical Merit *has sub-criteria* Importance*,* Contents *and* Depth*, and* Presentation *has sub-criteria* Style, Organization, Clarity *and* References*). Each reviewer scores the sub-criteria using a five-word vocabulary (*Poor, Marginal, Adequate, Good *and* Excellent*). After aggregation, the overall quality of the paper is represented as an IT2 FS. It can then be mapped into a word in the same five-word vocabulary for easy understanding.* ■

Example 3 *For the social judgment advisor introduced in [?, 46, 53], the CWW engine uses PR to infer an FOU representing the level of flirtation given the levels of touching and eye contact. The flirtation levels used in the consequents of the rules are from the 32-word vocabulary constructed in [15] [shown also in Fig. 4.1(a)]. To understand the CWW engine output, it is mapped into a word in the same 32-word vocabulary.* ■

Two previous approaches on the *Word* decoder are introduced next. Both of them focus on T1 FSs, but the ideas should also apply to IT2 FSs.

Yager [56] calls the *Word* decoder *Retranslation*, i.e., *"selecting a term from our prescribed vocabulary to express information represented using fuzzy sets,"* and he suggests four requirements that are involved in an objective[1] retranslation process:

1. *Validity* (or *Truth*), which is *"perhaps the most important requirement"* and is computed by an inclusion (subsethood) measure. Subsethood of FSs was first introduced by Zadeh [64] and then extended by Kosko [12], who defined the subsethood of a T1 FS A in another T1 FS B as

$$ss_K(A,B) = \frac{\int_{x\in X} \min(\mu_A(x), \mu_B(x))dx}{\int_{x\in X} \mu_A(x)dx} \tag{2.1}$$

or

$$ss_K(A,B) = \frac{\sum_{i=1}^{N} \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^{N} \mu_A(x_i)} \tag{2.2}$$

for the discrete case. Observe that $ss_K(A,B) \neq ss_K(B,A)$, and $ss_K(A,B) = 1$ if and only if $\mu_A(x_i) \leqslant \mu_B(x_i)$ for $\forall x_i$.

2. *Closeness* [10], which is related to the distance between the CWW engine output and its mapped word, because validity [56] *"does not necessarily imply closeness of fuzzy sets."*

3. *Fuzziness* [5, 49, 58, 59], which is [56] *"essentially related to the type of boundary distinguishing membership from nonmembership."*

4. *Specificity* [11, 60], which is [56] *"related to the quantity of information of certainty associated with a linguistic value."* Specificity is different from fuzziness because [56] *"whereas fuzziness is*

[1] Yager [56] also introduces a translation process that gives particular perceptions, e.g., *"a company trying to sell a financial product will try to express the information about the risk involved in the most favorable light;"* however, this kind of retranslation is out of the scope of this paper.

related to the type of boundary, the concept of specificity is related more closely to the cardinality or width of the set." For a normal T1 FS A, its specificity can be computed as [56]

$$Sp(A) = 1 - \frac{1}{|X|}\int_X \mu_A(x)dx \tag{2.3}$$

i.e., $Sp(A)$ is [56] *"the difference between the largest membership grade and the average of the remaining membership grades."*

Yager then proposes that an aggregation function, like the ordered weighted averages (OWAs) [55,57], can be used to aggregate these requirements in the retranslation process. No matter which aggregation function is used, the central idea here is that subsethood alone is not enough as a decoder when the desired recommendation is a word.

Martin and Klir [19] consider two criteria in the retranslation process:

1. *Validity*, which is again computed by Kosko's subsethood measure.
2. *Informativeness*, which is essentially the same as Yager's *specificity* requirement, except that it is computed by a different formula.

Martin and Klir [19] suggest a weighted average be used to combine these two criteria. No matter how the weights are determined, again, the central idea here is that subsethood alone is not enough as a decoder.

From the above discussion it seems that validity and specificity (informativeness) are the two most important criteria for a *Word* decoder. Interestingly, the Jaccard similarity measure for T1 FSs [8,51],

$$sm_J(A,B) = \frac{\sum_{i=1}^N \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^N \max(\mu_A(x_i), \mu_B(x_i))}; \tag{2.4}$$

combines these two criteria naturally: it closely resembles Kosko's subsethood measure [see (2.2)], but the denominator is changed from $\sum_{i=1}^N \mu_A(x_i)$ to $\sum_{i=1}^N \max(\mu_A(x_i), \mu_B(x_i))$, i.e., it also takes the width of A into consideration; so, it considers the specificity of A simultaneously, as illustrated by the following example. Note that this example uses T1 FSs; however, the arguments also hold for IT2 FSs.

Example 4 *Let A be the output of a CWW engine, and B and C be words in a pre-specified vocabulary, as shown in Fig. 2.1. It follows from Kosko's subsethood measure that $ss_K(A,B) = ss_K(A,C) = 1$, i.e., B and C are equally good; however, intuitively, since C is wider (less specific) than B, B should be chosen as the output. When the Jaccard similarity is used, we have $sm_J(A,B) > sm_J(A,C)$, which suggests that B is preferred over C.* ■

In conclusion, the Jaccard similarity measure can be viewed as a combination of a subsethood measure and a specificity measure, and hence it can be used in the word *decoder where the recommendation is a word from the same vocabulary used by the CWW engine.*

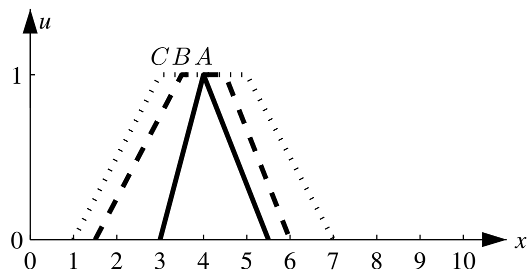

Fig. 2.1: A (solid curve), B (dashed curve) and C (dotted curve) used in Example 4.

2.3 Class

In some decision-making applications the output of the CWW engine needs to be mapped into a class or decision, whose FOU is stored in a *new* vocabulary which is different from the one used by the CWW engine.

Example 2 (Continued) *As pointed out in Example 2, the overall quality of a paper is aggregated as an IT2 FS, and it can be mapped into a word in the 5-word vocabulary (used by the CWW engine) for understanding. However, the usual goal is to classify this paper into one of three decision categories: Reject, Rewrite, or Accept. These categories represent a higher level of decision-making process than determining the quality of the paper, which is handled by the* word *decoder.* ■

Example 3 (Continued) *As pointed out in Example 3, the CWW engine outputs an FOU representing the level of flirtation given the levels of touching and eye contact, which can be mapped into a word in the same 32-word vocabulary used by the CWW engine for understanding. Suppose that the user also wants to take some actions based on the level of flirtation he/she receives, e.g.,* Consider it as harassment and report it to his/her supervisor*, or* do not consider it as harassment and ignore it*. These two actions can also be represented as IT2 FSs, but their spans would be wider than any other word FOUs in the 32-word codebook because each of these two decision categories may include many words in the 32-word vocabulary. Note that these two categories represent a higher level of decision-making process than determining the level of flirtation.* ■

From the above two examples we see that for the *Class* decoder, we need to determine the degree of containment of one set in another, which is the idea of subsethood. So, we propose that subsethood should be used as a decoder when the output recommendation is a class.

Chapter 3

Subsethood Measures for IT2 FSs

Three desirable properties for IT2 FS subsethood measure as a decoder and five existing IT2 FS subsethood measures are introduced in this section.

3.1 Desirable Properties for Subsethood Measure as a Decoder

Subsethood measures for FSs have been studied by many authors [3,6,16,32,35,38,41,61,63], and many different desirable properties or axioms have been proposed for them. In this paper we are interested in subsethood measures for IT2 FSs, particularly, their role as a decoder in the Per-C. Because the Jaccard similarity measure for T1 FSs and Kosko's subsethood measure for T1 FSs bare a strong resemblance, and the former is the basis for the Jaccard similarity measure for IT2 FSs [51] and the latter is the basis for almost all subsethood measures for IT2 FSs [3, 16, 32, 35, 41, 61], desirable properties for IT2 FS subsethood measures should be analogous to those for IT2 FS similarity measures [51]. Three desirable properties for IT2 FS subsethood measures are proposed in this subsection. But first, three definitions that are used in these desirable properties are introduced. These definitions have also been used in comparing IT2 FS similarity measures [31, 46, 51].

Definition 1 $\tilde{A} \leqslant \tilde{B}$ *if* $\overline{\mu}_{\tilde{A}}(x) \leqslant \overline{\mu}_{\tilde{B}}(x)$ *and* $\underline{\mu}_{\tilde{A}}(x) \leqslant \underline{\mu}_{\tilde{B}}(x)$ *for* $\forall x \in X$. ■

Examples of $\tilde{A} \leqslant \tilde{C} \leqslant \tilde{D}$ and $\tilde{B} \leqslant \tilde{C} \leqslant \tilde{D}$ are shown in Fig. 3.1.

Remarks: Note that our definition of $\tilde{A} \leqslant \tilde{B}$ is the same as the definition of $\tilde{A} \subseteq \tilde{B}$ in [3, 41], although we use $\leqslant$ instead of $\subseteq$. We avoid using $\subseteq$ because it is confusing, as some researchers [61] require $\tilde{A} \subseteq \tilde{B}$ to mean $\overline{\mu}_{\tilde{A}}(x) \leqslant \overline{\mu}_{\tilde{B}}(x)$ and $\underline{\mu}_{\tilde{A}}(x) \geqslant \underline{\mu}_{\tilde{B}}(x)$ for $\forall x \in X$, i.e., the FOU of $\tilde{A}$ is contained completely within the FOU of $\tilde{B}$. One such example is $\tilde{A}$ and $\tilde{B}$ in Fig. 3.1. Whether or not in this case $\tilde{A}$ should be considered completely as a subset of $\tilde{B}$ is an open problem.

Definition 2 $\tilde{A}$ *and* $\tilde{B}$ *overlap, i.e.,* $\tilde{A} \cap \tilde{B} \neq \emptyset$*, if and only if* $\exists x$ *such that* $\min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x)) > 0$. ■

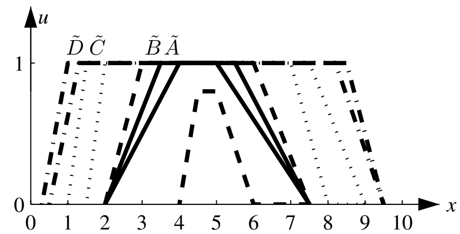

Fig. 3.1: $\tilde{A}$ (solid curves), $\tilde{B}$ (dashed curves), $\tilde{C}$ (dotted curve) and $\tilde{D}$ (dash-dotted curve) used in Example 6.

All IT2 FSs in Fig. 3.1 overlap with each other. Since only normal IT2 FSs are considered in the Per-C, overlapping $\tilde{A}$ and $\tilde{B}$ must intersect with each other at at least one point.

Definition 3 *$\tilde{A}$ and $\tilde{B}$ do not overlap, i.e., $\tilde{A} \cap \tilde{B} = \emptyset$, if and only if* $\min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x)) = \min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x)) = 0$ *for* $\forall x$. ■

Non-overlapping $\tilde{A}$ and $\tilde{B}$ have no parts of their FOUs that overlap, e.g., *None to very little* and *Maximum amount* in Fig. 4.1(a) do not overlap.

Definition 4 *Let $ss(\tilde{A}, \tilde{B})$ be a subsethood measure for IT2 FSs. Then, $ss(\tilde{A}, \tilde{B})$ should have the following desirable properties when it is used as a decoder in the Per-C[1]:*

1. *Reflexivity : $ss(\tilde{A}, \tilde{B}) = 1 \Leftrightarrow \tilde{A} \leqslant \tilde{B}$.*
2. *Transitivity: If $\tilde{C} \leqslant \tilde{A} \leqslant \tilde{B}$, then $ss(\tilde{A}, \tilde{C}) \geqslant ss(\tilde{B}, \tilde{C})$, and (more strongly) if $\tilde{A} \leqslant \tilde{B}$, then $ss(\tilde{C}, \tilde{A}) \leqslant ss(\tilde{C}, \tilde{B})$ for any $\tilde{C}$.*
3. *Overlapping: If $\tilde{A} \cap \tilde{B} \neq \emptyset$, then $ss(\tilde{A}, \tilde{B}) > 0$; otherwise, $ss(\tilde{A}, \tilde{B}) = 0$.* ■

Note that:

1. Reflexivity is a direct generalization of Zadeh's definition of containment [64] for T1 FSs, i.e., B contains A if $\mu_A(x) \leqslant \mu_B(x)$ for $\forall x \in X$. This property is also considered necessary for T1 FSs by Sinha and Dougherty [38], Young [63], Fan and Xie [6], and Zeng and Li [71], for intuitionistic FSs by Liu and Xiong [17] and Cornelis and Kerre [3], for general interval-valued intuitionistic FSs by Liu et al. [16], and for interval-valued FSs by Vlachos and Sergiadis [41]. As remarked under Definition 1, whether or not the subsethood of $\tilde{C}$ in $\tilde{B}$ in Fig. 3.1 should be 1 is an open problem. In this paper we require $ss(\tilde{C}, \tilde{B}) < 1$, as done by all authors [3,16,17,41] except Yang and Lin [61].
2. Transitivity is a direct generalization of Axiom (S3) for T1 FS subsethood measure proposed by Young [63], i.e., if $C \leqslant A \leqslant B$, then $ss(A, C) \geqslant ss(B, C)$, and if $A \leqslant B$, then $ss(C, A) \leqslant ss(C, B)$

[1] Some authors [3,32,36] define IT2 FS subsethood measures as intervals. For such cases, reflexivity becomes $ss(\tilde{A}, \tilde{B}) = [1, 1] \Leftrightarrow \tilde{A} \leqslant \tilde{B}$, and overlapping becomes: If $\tilde{A} \cap \tilde{B} \neq \emptyset$, then $ss(\tilde{A}, \tilde{B}) > [0, 0]$; otherwise, $ss(\tilde{A}, \tilde{B}) = [0, 0]$.

for any C. It has already been used for interval-valued FSs by Vlachos and Sergiadis [41]. The stronger transitivity for T1 FSs, i.e., if $A \leqslant B$ then $ss(A, C) \leqslant ss(B, C)$ and $ss(C, A) \leqslant ss(C, B)$ for any C, was proposed by Sinha and Dougherty [38]. The weaker transitivity, i.e., if $C \leqslant A \leqslant B$ then $ss(A, C) \leqslant ss(B, C)$ and $ss(C, A) \leqslant ss(C, B)$, was proposed by Fan et al. [6] and Zeng and Li [71] for T1 FSs, Liu and Xiong [17] for intuitionistic FSs, and Liu et al. [16] for general interval-valued intuitionistic FSs.

3. Overlapping is a direct generalization of Fan et al.'s [6] overlapping property for T1 FS subsethood measures, i.e., if[2] $A \neq 0$ and $A \cap B = 0$, then $ss(A, B) = 0$. We do not require $\tilde{A} \neq \emptyset$ in overlapping between $\tilde{A}$ and $\tilde{B}$ because this never happens in our applications.

4. For IT2 FS similarity measures there is another desirable property called symmetry, e.g., $sm_J(\tilde{A}, \tilde{B}) = sm_J(\tilde{B}, \tilde{A})$; however, generally IT2 FS subsethood measures are asymmetrical, i.e., $ss(\tilde{A}, \tilde{B}) \neq ss(\tilde{B}, \tilde{A})$. There are special conditions under which $ss(\tilde{A}, \tilde{B}) = ss(\tilde{B}, \tilde{A})$; however, usually these conditions are different for different subsethood measures, and it may be impossible to find an intuitive condition under which $ss(\tilde{A}, \tilde{B}) = ss(\tilde{B}, \tilde{A})$ should hold. So, symmetry is not considered in this paper.

Five existing IT2 FS subsethood measures are introduced next. They are compared in Section 4.

3.2 Liu and Xiong's Subsethood Measure for IT2 FSs

Liu and Xiong [17] proposed a subsethood measure for intuitionistic FSs [1]. Though intuitionistic FSs are generally different from IT2 FSs, Atanassov and Gargov [2] showed that every intuitionistic FS can be mapped into an interval valued FS [39], and is an IT2 FS under a different name. Using Atanassov and Gargov's mapping, Liu and Xiong's subsethood measure for IT2 FS is

$$ss_{LX}(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^{N} \min\left(1, 1 + \left(\underline{\mu}_{\tilde{B}}(x_i) + \overline{\mu}_{\tilde{B}}(x_i) - \underline{\mu}_{\tilde{A}}(x_i) - \overline{\mu}_{\tilde{A}}(x_i)\right)/2\right)}{N} \tag{3.1}$$

$ss_{LX}(\tilde{A}, \tilde{B})$ satisfies only transitivity but not reflexivity and overlapping. Counter-examples are shown in Section 4.

3.3 Cornelis and Kerre's Subsethood Measure for IT2 FSs

Cornelis and Kerre [3] introduced an interval inclusion (subsethood) measure for intuitionistic FSs. Using Atanassov and Gargov's mapping [2] from intuitionistic FSs into IT2 FSs, Cornelis and Kerre's

[2] $A \neq 0$ and $A \cap B = 0$ in [6] actually mean $A \neq \emptyset$ and $A \cap B = \emptyset$.

subsethood measure for IT2 FSs is

$$ss_{CK}(\tilde{A},\tilde{B}) = \left[\frac{\sum_{i=1}^{N}\min\Big(1, 1-\underline{\mu}_{\tilde{A}}(x_i)+\underline{\mu}_{\tilde{B}}(x_i), 1-\overline{\mu}_{\tilde{A}}(x_i)+\overline{\mu}_{\tilde{B}}(x_i)\Big)}{N}, \right.$$
$$\left. \frac{\sum_{i=1}^{N}\min\Big(1, 1-\underline{\mu}_{\tilde{A}}(x_i)+\overline{\mu}_{\tilde{B}}(x_i)\Big)}{N}\right] \tag{3.2}$$

$ss_{CK}(\tilde{A},\tilde{B})$ satisfies transitivity and reflexivity but not overlapping. Counter-examples are shown in Section 4.

3.4 Vlachos and Sergiadis's Subsethood Measure for IT2 FSs

Vlachos and Sergiadis [41] proposed a subsethood measure for interval-valued FSs, which are the same as IT2 FSs. It is

$$ss_{VS}(\tilde{A},\tilde{B}) = 1 - \frac{\sum_{i=1}^{N}\max(0,\underline{\mu}_{\tilde{A}}(x_i)-\underline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N}\max(0,\overline{\mu}_{\tilde{A}}(x_i)-\overline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{N}\underline{\mu}_{\tilde{A}}(x_i)+\overline{\mu}_{\tilde{A}}(x_i)} \tag{3.3}$$

It will be shown in Section 4 that $ss_{VS}(\tilde{A},\tilde{B})$ is the best IT2 FS subsethood measure to use as a decoder in the Per-C; hence, more details about it are presented next.

When $\underline{\mu}_{\tilde{A}}(x_i) \geqslant \underline{\mu}_{\tilde{B}}(x_i)$, it follows that

$$\begin{aligned}\underline{\mu}_{\tilde{A}}(x_i) - \max(0,\underline{\mu}_{\tilde{A}}(x_i)-\underline{\mu}_{\tilde{B}}(x_i)) &= \underline{\mu}_{\tilde{A}}(x_i) - \underline{\mu}_{\tilde{A}}(x_i) + \underline{\mu}_{\tilde{B}}(x_i)\\ &= \underline{\mu}_{\tilde{B}}(x_i) = \min(\underline{\mu}_{\tilde{A}}(x_i),\underline{\mu}_{\tilde{B}}(x_i))\end{aligned} \tag{3.4}$$

and when $\underline{\mu}_{\tilde{A}}(x_i) < \underline{\mu}_{\tilde{B}}(x_i)$, it follows that

$$\underline{\mu}_{\tilde{A}}(x_i) - \max(0,\underline{\mu}_{\tilde{A}}(x_i)-\underline{\mu}_{\tilde{B}}(x_i)) = \underline{\mu}_{\tilde{A}}(x_i) = \min(\underline{\mu}_{\tilde{A}}(x_i),\underline{\mu}_{\tilde{B}}(x_i)); \tag{3.5}$$

hence,

$$\underline{\mu}_{\tilde{A}}(x_i) - \max(0,\underline{\mu}_{\tilde{A}}(x_i)-\underline{\mu}_{\tilde{B}}(x_i)) = \min(\underline{\mu}_{\tilde{A}}(x_i),\underline{\mu}_{\tilde{B}}(x_i)) \tag{3.6}$$

Similarly,

$$\overline{\mu}_{\tilde{A}}(x_i) - \max(0,\overline{\mu}_{\tilde{A}}(x_i)-\overline{\mu}_{\tilde{B}}(x_i)) = \min(\overline{\mu}_{\tilde{A}}(x_i),\overline{\mu}_{\tilde{B}}(x_i)) \tag{3.7}$$

Consequently, (3.3) can be re-written as

$$\begin{aligned}ss_{VS}(\tilde{A},\tilde{B}) &= \frac{\sum_{i=1}^{N}[\underline{\mu}_{\tilde{A}}(x_i) - \max(0,\underline{\mu}_{\tilde{A}}(x_i)-\underline{\mu}_{\tilde{B}}(x_i))] + \sum_{i=1}^{N}[\overline{\mu}_{\tilde{A}}(x_i) - \max(0,\overline{\mu}_{\tilde{A}}(x_i)-\overline{\mu}_{\tilde{B}}(x_i))]}{\sum_{i=1}^{N}\underline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^{N}\overline{\mu}_{\tilde{A}}(x_i)}\\ &= \frac{\sum_{i=1}^{N}\min(\underline{\mu}_{\tilde{A}}(x_i),\underline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N}\min(\overline{\mu}_{\tilde{A}}(x_i),\overline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{N}\underline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^{N}\overline{\mu}_{\tilde{A}}(x_i)}\end{aligned} \tag{3.8}$$

Observe the analogy of (3.8) to the Jaccard similarity measure in (1.5). In fact, we can further re-express (3.8) as

$$ss_{VS}(\tilde{A},\tilde{B}) = \frac{\frac{|X|}{2N}[\sum_{i=1}^{N}\min(\underline{\mu}_{\tilde{A}}(x_i),\underline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N}\min(\overline{\mu}_{\tilde{A}}(x_i),\overline{\mu}_{\tilde{B}}(x_i))]}{\frac{|X|}{2N}[\sum_{i=1}^{N}\underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=1}^{N}\overline{\mu}_{\tilde{A}}(x_i)]} \tag{3.9}$$

Observe that the numerator of (3.10) is the average cardinality [46,49] of $\tilde{A}\cap\tilde{B}$, and the denominator of (3.10) is the average cardinality of $\tilde{A}$. This is similar to the way the Jaccard similarity measure for IT2 FSs was developed in [51], i.e., the numerator of (1.5) is the average cardinality of $\tilde{A}\cap\tilde{B}$, and the denominator of (1.5) is the average cardinality of $\tilde{A}\cup\tilde{B}$.

Since (3.8) is easier to understand and to compute than (3.3), it is used in the rest of the paper instead of (3.3).

The continuous version of Vlachos and Sergiadis's IT2 FS subsethood measure can be expressed as

$$ss_{VS}(\tilde{A},\tilde{B}) = \frac{\int_X \min(\underline{\mu}_{\tilde{A}}(x),\underline{\mu}_{\tilde{B}}(x))dx + \int_X \min(\overline{\mu}_{\tilde{A}}(x),\overline{\mu}_{\tilde{B}}(x))dx}{\int \underline{\mu}_{\tilde{A}}(x)dx + \int \overline{\mu}_{\tilde{A}}(x)dx} \tag{3.10}$$

The following theorem describes three properties of $ss_{VS}(\tilde{A},\tilde{B})$; its proof is given in Appendix A. Illustrative examples are given in Section 4.

Theorem 1 $ss_{VS}(\tilde{A},\tilde{B})$ *satisfies transitivity, reflexivity and overlapping.* ■

3.5 Rickard et al.'s Subsethood Measure for IT2 FSs

Rickard et al. [35, 36] and Nguyen and Kreinovich [32] independently extended Kosko's subsethood measure to IT2 FSs based on the Representation Theorem [28], which states that an IT2 FS $\tilde{A}$ is the union of all of its embedded T1 FSs. Because Rickard et al.'s results were published a little earlier than Nguyen and Kreinovich's, in this paper we call this subsethood measure Rickard et al.'s subsethood measure, which is

$$ss_R(\tilde{A},\tilde{B}) \equiv [ss_l(\tilde{A},\tilde{B}), ss_r(\tilde{A},\tilde{B})] \tag{3.11}$$

where

$$ss_l(\tilde{A},\tilde{B}) = \min_{\mu_{A_l}(x_i) \text{ in } (3.14)} \left[\frac{\sum_{i=1}^{N}\min\left(\mu_{A_l}(x_i),\underline{\mu}_{\tilde{B}}(x_i)\right)}{\sum_{i=1}^{N}\mu_{A_l}(x_i)}\right] \tag{3.12}$$

$$ss_r(\tilde{A},\tilde{B}) = \frac{\sum_{i=1}^{N}\min\left(\mu_{A_r}(x_i),\overline{\mu}_{\tilde{B}}(x_i)\right)}{\sum_{i=1}^{N}\mu_{A_r}(x_i)} \tag{3.13}$$

in which

$$\mu_{A_l}(x_i) = \begin{cases} \overline{\mu}_{\tilde{A}}(x_i), & \underline{\mu}_{\tilde{B}}(x_i) \leqslant \underline{\mu}_{\tilde{A}}(x_i) \\ \underline{\mu}_{\tilde{A}}(x_i), & \underline{\mu}_{\tilde{B}}(x_i) \geqslant \overline{\mu}_{\tilde{A}}(x_i) \\ \underline{\mu}_{\tilde{A}}(x_i) \text{ or } \overline{\mu}_{\tilde{A}}(x_i), & x_i \in I_l \end{cases} \tag{3.14}$$

$$\mu_{A_r}(x_i) = \begin{cases} \underline{\mu}_{\tilde{A}}(x_i), & \overline{\mu}_{\tilde{B}}(x_i) \leqslant \underline{\mu}_{\tilde{A}}(x_i) \\ \overline{\mu}_{\tilde{A}}(x_i), & \overline{\mu}_{\tilde{B}}(x_i) \geqslant \overline{\mu}_{\tilde{A}}(x_i) \\ \overline{\mu}_{\tilde{B}}(x_i), & \underline{\mu}_{\tilde{A}}(x_i) < \overline{\mu}_{\tilde{B}}(x_i) < \overline{\mu}_{\tilde{A}}(x_i) \end{cases} \tag{3.15}$$

and

$$I_l \equiv \{x_i | \underline{\mu}_{\tilde{A}}(x_i) < \underline{\mu}_{\tilde{B}}(x_i) < \overline{\mu}_{\tilde{A}}(x_i)\} \tag{3.16}$$

Note that $ss_r(\tilde{A}, \tilde{B})$ has a closed-form solution; however, because for each $x_i \in I_l$, $\mu_{A_l}(x_i)$ can have two possible values, to compute $ss_l(\tilde{A}, \tilde{B})$, 2^L evaluations of the bracketed term in (3.12) have to be performed, where L is the number of elements in I_l, and this can be a rather large number depending upon L. An efficient algorithm for computing $ss_l(\tilde{A}, \tilde{B})$ is proposed in [32].

$ss_R(\tilde{A}, \tilde{B})$ satisfies transitivity and overlapping but not reflexivity. Counter-examples are shown in Section 4.

3.6 Yang and Lin's Subsethood Measure for T2 FSs

Yang and Lin [61] defined an inclusion (subsethood) measure for general T2 FSs [22, 27]. When specialized to IT2 FSs, it becomes

$$ss_{YL}(\tilde{A}, \tilde{B}) = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{u_j \in \theta_i} u_j}{\sum_{u_j \in \varphi_i} u_j} \tag{3.17}$$

where

$$\theta_i = [\max(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i)), \min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i))] \tag{3.18}$$

$$\varphi_i = [\underline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{A}}(x_i)] \tag{3.19}$$

$ss_{YL}(\tilde{A}, \tilde{B})$ satisfies none of reflexivity[3], transitivity and overlapping. Counter-examples are shown in Section 4.

[3]Yang and Lin [61] proved that $\tilde{A} \subseteq \tilde{B} \Leftrightarrow ss_{YL}(\tilde{A}, \tilde{B}) = 1$ (Property 1 in [61]), but their definition of $\tilde{A} \subseteq \tilde{B}$ for IT2 FSs is $\underline{\mu}_{\tilde{A}}(x) \geqslant \underline{\mu}_{\tilde{B}}(x)$ and $\overline{\mu}_{\tilde{A}}(x) \leqslant \overline{\mu}_{\tilde{B}}(x)$ for $\forall x \in X$, which is different from the definition of $\tilde{A} \leqslant \tilde{B}$ in Definition 1 of this paper. Yang and Lin also proved some nice properties of $ss_{YL}(\tilde{A}, \tilde{B})$ (Property 5 in [61]), e.g., $ss_{YL}(\tilde{A} \cup \tilde{B}, \tilde{C}) = \min\{ss_{YL}(\tilde{A}, \tilde{C}), ss_{YL}(\tilde{B}, \tilde{C})\}$, $ss_{YL}(\tilde{A} \cap \tilde{B}, \tilde{C}) = \max\{ss_{YL}(\tilde{A}, \tilde{C}), ss_{YL}(\tilde{B}, \tilde{C})\}$, etc; however, their proofs were incomplete because they only considered the cases that $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are included in each other, e.g., $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$, $\tilde{B} \subseteq \tilde{A} \subseteq \tilde{C}$, etc. Cases like $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ shown in Fig. 3.1, where $\tilde{A} \nsubseteq \tilde{B}$, $\tilde{A} \nsubseteq \tilde{C}$, $\tilde{B} \nsubseteq \tilde{A}$ and $\tilde{B} \nsubseteq \tilde{C}$, were not considered. In general, these properties of $ss_{YL}(\tilde{A}, \tilde{B})$ are not true.

Chapter 4

Comparative Studies

In this section several examples are used to compare the five subsethood measures.

Example 6 *In this example, $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ and $\tilde{D}$ shown in Fig. 3.1 are used to compare the five subsethood measures. Results are given in Table 4.1. Some highlights of Table 4.1 are explained in the following:*

1. *According to reflexivity, we should have $ss(\tilde{A}, \tilde{A}) = 1$, or $ss(\tilde{A}, \tilde{A}) = [1, 1]$ if $ss(\tilde{A}, \tilde{A})$ is an interval subsethood measure; however, Table 4.1 shows that $ss_R(\tilde{A}, \tilde{A}) \neq [1, 1]$, i.e., ss_R does not satisfy reflexivity.*

2. *Since $\tilde{B} \nleq \tilde{A}$ and $\tilde{A} \nleq \tilde{B}$, it follows from reflexivity that $ss(\tilde{B}, \tilde{A}) \neq 1$; however, observe from Table 4.1 that $ss_{LX}(\tilde{B}, \tilde{A}) = 1$, i.e., ss_{LX} does not satisfy reflexivity.*

3. *Since $\tilde{C} \leqslant \tilde{D}$, from reflexivity it follows that $ss(\tilde{C}, \tilde{D}) = 1$; however, Table 4.1 shows that $ss_{YL}(\tilde{C}, \tilde{D}) < 1$, i.e., ss_{YL} does not satisfy reflexivity.*

4. *Since $\tilde{C} \leqslant \tilde{D}$, from transitivity it follows that $ss(\tilde{C}, \tilde{D}) \geqslant ss(\tilde{C}, \tilde{C}) = 1$, i.e., $ss(\tilde{C}, \tilde{D}) = 1$; however, Table 4.1 shows that $ss_{YL}(\tilde{C}, \tilde{D}) < ss_{YL}(\tilde{C}, \tilde{C}) < 1$, i.e., ss_{YL} does not satisfy transitivity.*

5. *Only ss_{CK} and ss_{VS} give results that, in all cases, satisfy reflexivity and transitivity.* ■

The following example illustrates the overlapping property, and also explains why the Jaccard similarity measure should not be used as a classifier in decoding.

Example 7 *In this example, the 32 word FOUs obtained using the IA of [15] [see Fig. 4.1(a)] are used to compare the differences among the five subsethood measures and the Jaccard similarity measure. For simplicity, we define two classes in the [0, 10] interval [see Fig. 4.1(b)]:*

$$\tilde{B} = \textit{Smaller than or equal to 5} \tag{4.1}$$

$$\tilde{C} = \textit{Larger than 5} \tag{4.2}$$

Table 4.1: Comparison of the five subsethood measures for $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ and $\tilde{D}$ in Fig. 3.1. Results that are not consistent with the desirable results derived from the three desirable subsethood properties are marked in bold.

		Desirable results	ss_{LX}	ss_{CK}	ss_{VS}	ss_{R}	ss_{YL}
$\tilde{A}$	$\tilde{A}$	$ss(\tilde{A},\tilde{A})=1$	1	[1, 1]	1	**[0.87, 1]**	1
$\tilde{A}$	$\tilde{B}$	$ss(\tilde{A},\tilde{B})<1$	0.84	[0.59,1]	0.68	[0.27, 1]	**1**
$\tilde{A}$	$\tilde{C}$	$ss(\tilde{A},\tilde{C})=1$	1	[1, 1]	1	[1, 1]	**0.18**
$\tilde{A}$	$\tilde{D}$	$ss(\tilde{A},\tilde{D})=1$	1	[1, 1]	1	[1, 1]	**0.18**
$\tilde{B}$	$\tilde{A}$	$ss(\tilde{B},\tilde{A})<1$	**1**	[.91, 1]	0.90	[0.72, 1]	0.20
$\tilde{B}$	$\tilde{B}$	$ss(\tilde{B},\tilde{B})=1$	1	[1, 1]	1	**[0.24, 1]**	1
$\tilde{B}$	$\tilde{C}$	$ss(\tilde{B},\tilde{C})=1$	1	[1, 1]	1	[1, 1]	**0.01**
$\tilde{B}$	$\tilde{D}$	$ss(\tilde{B},\tilde{D})=1$	1	[1, 1]	1	[1, 1]	**0.01**
$\tilde{C}$	$\tilde{A}$	$ss(\tilde{C},\tilde{A})<1$	0.64	[0.54, 0.75]	0.55	[0.46, 0.65]	0.26
$\tilde{C}$	$\tilde{B}$	$ss(\tilde{C},\tilde{B})<1$	0.53	[0.26, 0.81]	0.41	[0.14, 0.74]	0.39
$\tilde{C}$	$\tilde{C}$	$ss(\tilde{C},\tilde{C})=1$	1	[1, 1]	1	**[0.82, 1]**	1
$\tilde{C}$	$\tilde{D}$	$ss(\tilde{C},\tilde{D})=1$; $ss(\tilde{C},\tilde{D})\geqslant ss(\tilde{C},\tilde{C})$	1	[1, 1]	1	[1, 1]	**0.63**
$\tilde{D}$	$\tilde{A}$	$ss(\tilde{D},\tilde{A})\leqslant ss(\tilde{D},\tilde{C})$	0.50	[0.45, 0.55]	0.43	[0.39, 0.47]	0.23
$\tilde{D}$	$\tilde{B}$	$ss(\tilde{D},\tilde{B})\leqslant ss(\tilde{D},\tilde{C})$	0.40	[0.21, 0.60]	0.32	[0.12, 0.53]	0.34
$\tilde{D}$	$\tilde{C}$	$ss(\tilde{D},\tilde{C})\geqslant ss(\tilde{D},\tilde{A})$	0.81	[0.73, 0.90]	0.78	[0.69, 0.88]	0.66
$\tilde{D}$	$\tilde{D}$	$ss(\tilde{D},\tilde{D})=1$	1	[1, 1]	1	**[0.95, 1]**	1

Observe that $\tilde{B}$ and $\tilde{C}$ are actually crisp sets; however, they can be viewed as special IT2 FSs, e.g., $\tilde{B}$ has a normal trapezoidal UMF $[0,0,5,5]$ and a normal trapezoidal LMF $[0,0,5,5]$. By viewing the FOUs in Fig. 4.1(a) in relation to the dashed line drawn at $x=5$, it is clear that Words 1-10 are definitely in $\tilde{B}$ and Words 11-32 are definitely in $\tilde{C}$. The five subsethood measures of these 32 words in $\tilde{B}$ and $\tilde{C}$, as well as their Jaccard similarities with $\tilde{B}$ and $\tilde{C}$, are summarized in Table 4.2.

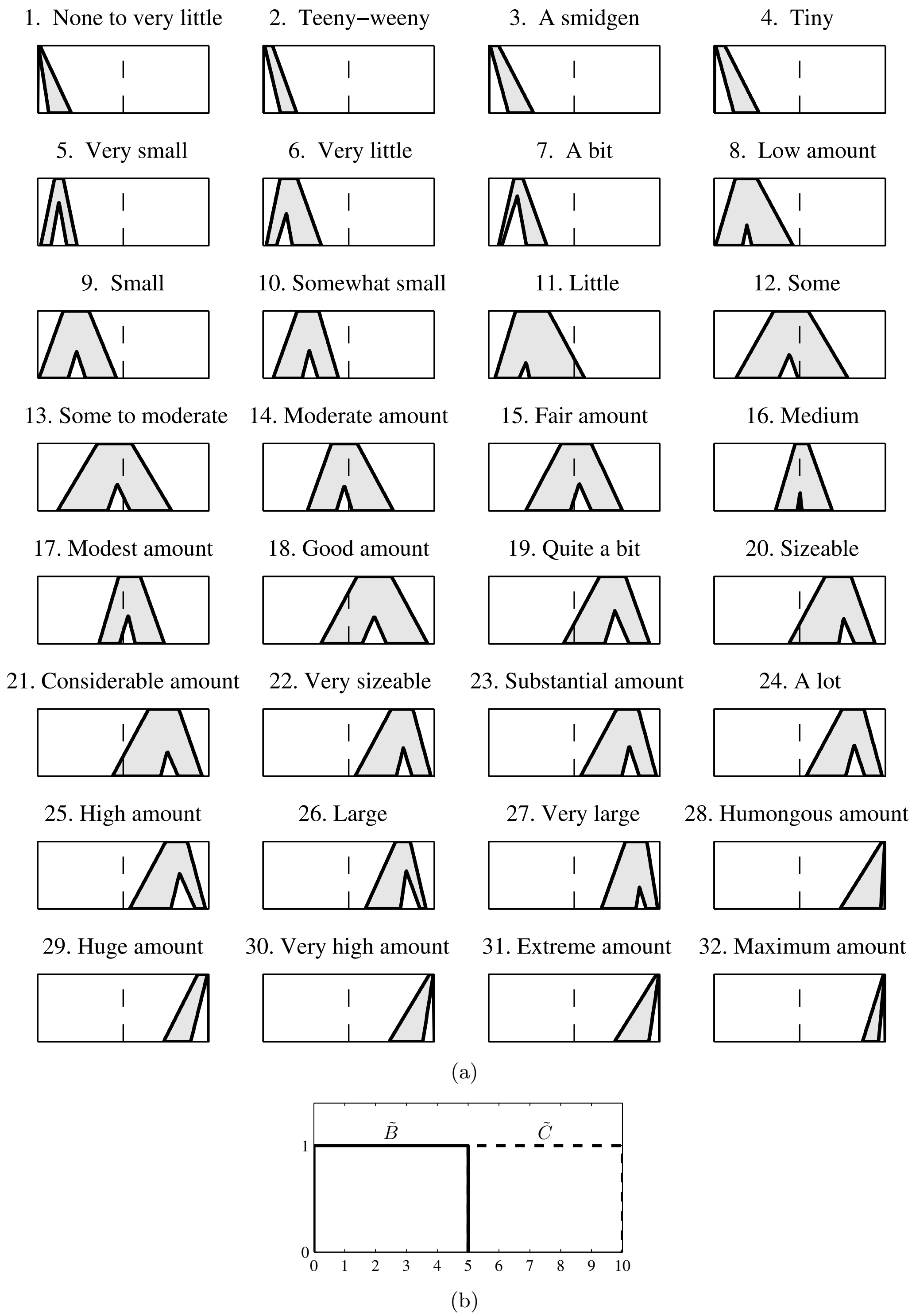

Fig. 4.1: Words and their FOUs used in Example 7. (a) The 32 words ranked by the centroid-based ranking method [51]. The dashed line is $x = 5$. (b) $\tilde{B}$ = *Smaller than or equal to 5* (solid curve) and $\tilde{C}$ = *Larger than 5* (dashed curve).

Table 4.2: Comparison of the five IT2 FS subsethood measures and the Jaccard similarity measure for the 32 FOUs in Fig. 4.1(a).

$\tilde{A}$	$\tilde{B}$ = Smaller than or equal to 5						$\tilde{C}$ = Larger than 5					
	ss_{LX}	ss_{CK}	ss_{VS}	ss_R	ss_{YL}	ss_J	ss_{LX}	ss_{CK}	ss_{VS}	ss_R	ss_{YL}	ss_J
1. None to very little	1	[1, 1]	1	[1, 1]	0.03	0.14	0.93	[0.90, 0.97]	0	[0, 0]	0.01	0
2. Teeny-weeny	1	[1, 1]	1	[1, 1]	0.05	0.18	0.91	[0.88, 0.95]	0	[0, 0]	0.01	0
3. A smidgen	1	[1, 1]	1	[1, 1]	0.04	0.22	0.89	[0.84, 0.94]	0	[0, 0]	0.01	0
4. Tiny	1	[1, 1]	1	[1, 1]	0.04	0.22	0.89	[0.84, 0.94]	0	[0, 0]	0.01	0
5. Very small	1	[1, 1]	1	[1, 1]	0	0.16	0.92	[0.87, 0.97]	0	[0, 0]	0.01	0
6. Very little	1	[1, 1]	1	[1, 1]	0	0.23	0.88	[0.78, 0.98]	0	[0, 0]	0.01	0
7. A bit	1	[1, 1]	1	[1, 1]	0	0.22	0.88	[0.82, 0.94]	0	[0, 0]	0.01	0
8. Low amount	1	[1, 1]	1	[1, 1]	0	0.30	0.85	[0.71, 0.99]	0	[0, 0]	0.01	0
9. Small	1	[1, 1]	1	[1, 1]	0.01	0.32	0.84	[0.70, 0.98]	0	[0, 0]	0.01	0
10. Somewhat small	1	[1, 1]	1	[1, 1]	0.01	0.27	0.86	[0.73, 0.98]	0	[0, 0]	0.01	0
11. Little	0.99	[0.98, 1]	0.98	[0.46, 1]	0.01	0.35	0.82	[0.64, 0.99]	0.03	[0, 0.56]	0.01	0.01
12. Some	0.90	[0.79,1]	0.63	[0.11, 1]	0.01	0.24	0.84	[0.70, 0.98]	0.38	[0, 0.89]	0.01	0.13
13. Some to moderate	0.89	[0.79, 0.99]	0.63	[0.12, 0.99]	0.01	0.25	0.84	[0.70, 0.98]	0.37	[0.02, 0.88]	0.01	0.13
14. Moderate amount	0.90	[0.80, 1]	0.54	[0.09, 0.99]	0.01	0.16	0.88	[0.77, 0.98]	0.46	[0.01, 0.91]	0.01	0.13
15. Fair amount	0.86	[0.76, 0.97]	0.46	[0.01, 0.89]	0.01	0.15	0.89	[0.77, 1]	0.55	[0.12, 0.99]	0.01	0.18
16. Medium	0.91	[0.82, 1]	0.40	[0.01, 0.97]	0.01	0.07	0.94	[0.87, 1]	0.60	[0.03, 0.99]	0.01	0.11
17. Modest amount	0.87	[0.77, 0.98]	0.31	[0.01, 0.83]	0.01	0.07	0.93	[0.87, 1]	0.69	[0.18, 0.99]	0.01	0.17
18. Good amount	0.80	[0.64, 0.97]	0.14	[0, 0.68]	0.01	0.05	0.95	[0.91, 1]	0.86	[0.32, 1]	0.01	0.36
19. Quite a bit	0.81	[0.66, 0.96]	0.02	[0, 0.20]	0.01	0.01	0.99	[0.98, 1]	0.98	[0.80, 1]	0.01	0.35
20. Sizeable	0.82	[0.66, 0.98]	0.03	[0, 0.33]	0.01	0.01	0.99	[0.98, 1]	0.97	[0.67, 1]	0.01	0.33
21. Considerable amount	0.81	[0.65, 0.98]	0.02	[0, 0.33]	0.01	0.01	0.99	[0.98, 1]	0.98	[0.68, 1]	0.01	0.36
22. Very sizeable	0.85	[0.71, 0.98]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0	0.30
23. Substantial amount	0.84	[0.71, 0.98]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0	0.31
24. A lot	0.84	[0.71, 0.97]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0	0.31
25. High amount	0.84	[0.71, 0.96]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0	0.32
26. Large	0.87	[0.77, 0.97]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0	0.25
27. Very large	0.88	[0.77, 0.99]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0.01	0.24
28. Humongous amount	0.93	[0.86, 0.99]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0.01	0.15
29. Huge amount	0.89	[0.84, 0.95]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0.03	0.22
30. Very high amount	0.91	[0.86, 0.97]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0.02	0.18
31. Extreme amount	0.92	[0.86, 0.97]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0.02	0.17
32. Maximum amount	0.96	[0.93, 0.98]	0	[0, 0]	0.01	0	1	[1, 1]	1	[1, 1]	0.03	0.09

Observe from Fig. 4.1(a) and Table 4.2 that:

1. *According to reflexivity, Words 1-10 should have* $ss(\tilde{A},\tilde{B})=1$ *since each of them is completely under or on* $\tilde{B}$*, and Words 11-32 should have* $ss(\tilde{A},\tilde{B})<1$ *since none of them is completely under or on* $\tilde{B}$*.* ss_{LX}*,* ss_{CK}*,* ss_{VS} *and* ss_R *satisfy this requirement, but* ss_{YL} *does not. Similarly, Words 1-21 should have* $ss(\tilde{A},\tilde{C})<1$ *and Words 22-32 should have* $ss(\tilde{A},\tilde{C})=1$*. Again,* ss_{LX}*,* ss_{CK}*,* ss_{VS} *and* ss_R *satisfy this requirement, but* ss_{YL} *does not.*

2. *According to overlapping, Words 1-21 should have* $ss(\tilde{A},\tilde{B})>0$ *since they overlap with* $\tilde{B}$*, and Words 22-32 should have* $ss(\tilde{A},\tilde{B})=0$ *since they do not overlap with* $\tilde{B}$*.* ss_{VS} *and* ss_R *satisfy this requirement, but* ss_{LX}*,* ss_{CK} *and* ss_{YL} *do not. Similarly, Words 1-10 should have* $ss(\tilde{A},\tilde{C})=0$ *and Words 11-32 should have* $ss(\tilde{A},\tilde{C})>0$*. Again,* ss_{VS} *and* ss_R *satisfy this requirement, but* ss_{LX}*,* ss_{CK} *and* ss_{YL} *do not.*

3. *As the word* $\tilde{A}$ *gets larger-sounding, generally* $ss_{VS}(\tilde{A},\tilde{B})$ *and* $ss_R(\tilde{A},\tilde{B})$ *decrease whereas* $ss_{VS}(\tilde{A},\tilde{C})$ *and* $ss_R(\tilde{A},\tilde{C})$ *increase, both of which are reasonable.*

4. *As shown in Table 4.2, for all words*[1] $\tilde{A}$*,*

$$ss_{VS}(\tilde{A},\tilde{B})+ss_{VS}(\tilde{A},\tilde{C})=1 \tag{4.3}$$

This relationship is always true for complementary crisp sets, e.g., $\tilde{B}$ *and* $\tilde{C}$ *in Fig. 4.1(b), because*

$$\begin{aligned}
ss_{VS}(\tilde{A},\tilde{B})+ss_{VS}(\tilde{A},\tilde{C}) =&\frac{\sum_{x_i\leqslant 5}\min(\overline{\mu}_{\tilde{A}}(x_i),\overline{\mu}_{\tilde{B}}(x_i))+\sum_{x_i\leqslant 5}\min(\underline{\mu}_{\tilde{A}}(x_i),\underline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}\\
&+\frac{\sum_{x_i>5}\min(\overline{\mu}_{\tilde{A}}(x_i),\overline{\mu}_{\tilde{C}}(x_i))+\sum_{x_i>5}\min(\underline{\mu}_{\tilde{A}}(x_i),\underline{\mu}_{\tilde{C}}(x_i))}{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}\\
=&\frac{\sum_{x_i\leqslant 5}\min(\overline{\mu}_{\tilde{A}}(x_i),1)+\sum_{x_i\leqslant 5}\min(\underline{\mu}_{\tilde{A}}(x_i),1)}{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}\\
&+\frac{\sum_{x_i>5}\min(\overline{\mu}_{\tilde{A}}(x_i),1)+\sum_{x_i>5}\min(\underline{\mu}_{\tilde{A}}(x_i),1)}{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}\\
=&\frac{\sum_{x_i\leqslant 5}\overline{\mu}_{\tilde{A}}(x_i)+\sum_{x_i\leqslant 5}\underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}+\frac{\sum_{x_i>5}\overline{\mu}_{\tilde{A}}(x_i)+\sum_{x_i>5}\underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}\\
=&\frac{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^N\overline{\mu}_{\tilde{A}}(x_i)+\sum_{i=1}^N\underline{\mu}_{\tilde{A}}(x_i)}\\
=&1.
\end{aligned} \tag{4.4}$$

5. *The Jaccard similarity measure cannot be used as a classifier because it gives counter-intuitive results, e.g.,* None to very little *should belong to the class* Smaller than or equal to 5 *completely, but its Jaccard similarity is only 0.14. Furthermore, as mentioned in Section 2.3, a classifier is needed when the output of the CWW engine,* $\tilde{A}$*, is mapped into a decision category* $\tilde{B}$*. Subsethood*

[1] In Table 4.2 sometimes (4.3) does not hold exactly because of discretization and roundoff errors.

is conceptually more appropriate for a classifier, because it defines the degree that $\tilde{A}$ is contained in $\tilde{B}$. On the other hand, it is not reasonable to compare the similarity between $\tilde{A}$ and $\tilde{B}$ because they belong to different domains (vocabularies), e.g., in Example 2 $\tilde{A}$ represents the overall quality of a paper, whereas $\tilde{B}$ is a publication decision. ■

Remarks: Note that our argument that similarity should not be used as a classifier in decoding for the Per-C does not mean that similarity should never be used for classification [33]. In some classification problems [4, 20], each training example has its associated class label, but the unique features of each class (i.e., the counterparts of the class-FOUs in decoding) are unknown. In this case, a similarity-based classifier can be constructed as follows: for each new input, the training example most similar to it is found, and the new input is then assigned to that training example's class. Subsethood-based classifier is preferred in the Per-C because the decoding problem considered in the Per-C is different, i.e., the class-FOUs ($\tilde{B}$) are known ahead of time and are not in the same domain as $\tilde{A}$.

Since T1 FSs are special cases of IT2 FSs, the five subsethood measures should also be applicable to T1 FSs. The following example illustrates this.

Example 8 *This example is used to compare the performances of the five subsethood measures when IT2 FSs degrade to T1 FSs, as shown in Fig. 4.2. The results are given in Table 4.3, where the last column of Table 4.3 shows the results obtained from Kosko's subsethood measure for T1 FSs, which is the* de facto *subsethood measure for T1 FSs. Observe that only ss_{VS} and ss_R are consistent with Kosko's subsethood measure.* ■

Table 4.3: Comparison of the five subsethood measures for A, B and C in Fig. 4.2.

		ss_{LX}	ss_{CK}	ss_{VS}	ss_R	ss_{YL}	ss_K
A	A	1	[1, 1]	1	[1, 1]	1	1
A	B	0.47	[0.47, 0.47]	0.32	[0.32, 0.32]	0.01	0.32
A	C	0.64	[0.64, 0.64]	0.55	[0.55, 0.55]	0.15	0.55
B	A	1	[1, 1]	1	[1, 1]	0	1
B	B	1	[1, 1]	1	[1, 1]	1	1
B	C	0.89	[0.89, 0.89]	0.64	[0.64, 0.64]	0.01	0.64
C	A	1	[1, 1]	1	[1, 1]	0.20	1
C	B	0.66	[0.66, 0.66]	0.38	[0.38, 0.38]	0.01	0.38
C	C	1	[1, 1]	1	[1, 1]	1	1

Table 4.4: Summarization of the properties of the five IT2 FS subsethood measures.

Property	ss_{LX}	ss_{CK}	ss_{VS}	ss_R	ss_{YL}
Reflexivity	×	✓	✓	×	×
Transitivity	✓	✓	✓	✓	×
Overlapping	×	×	✓	✓	×
Consistency with ss_K	×	×	✓	✓	×

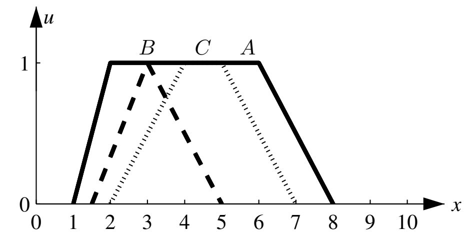

Fig. 4.2: A (solid curves), B (dashed curves) and C (dotted curves) used in Example 8.

Finally, Table 4.4 summarizes the properties of the five IT2 FS subsethood measures. It is clear that ss_{VS} is the best IT2 FS subsethood measure for the Per-C, because it satisfies all three desirable properties and is consistent with ss_K for T1 FSs. ss_R is the second best, and its only drawback is that usually $ss_R(\tilde{A}, \tilde{B}) \neq [1, 1]$ when $\tilde{A} = \tilde{B}$.

Chapter 5

Conclusions

This paper has explained the need for subsethood measures as a decoder in the Per-C, introduced three desirable properties for IT2 FS subsethood measures, and compared five existing IT2 FS subsethood measures against these properties. Examples have shown that when the desired recommendation is a class, Vlachos and Sergiadis's IT2 FS subsethood measure gives the most reasonable results. The results in this paper will be useful in constructing a third kind of decoder (i.e., in addition to similarity measures and ranking methods) for perceptual computing.

Appendix A

Proof of Theorem 1

Our proof of Theorem 1 is for the discrete case (3.8). The proof for the continuous case (3.10) is very similar, and is left to the reader.

1. *Reflexivity:* When $ss_{VS}(\tilde{A}, \tilde{B}) = 1$, it follows from (3.8) that,

$$\min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i)) = \underline{\mu}_{\tilde{A}}(x_i) \quad \forall x_i \tag{A.0-1}$$

$$\min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) = \overline{\mu}_{\tilde{A}}(x_i) \quad \forall x_i \tag{A.0-2}$$

i.e., $\overline{\mu}_{\tilde{B}}(x_i) \geqslant \overline{\mu}_{\tilde{A}}(x_i)$ and $\underline{\mu}_{\tilde{B}}(x_i) \geqslant \underline{\mu}_{\tilde{A}}(x_i)$; hence, $\tilde{A} \leqslant \tilde{B}$ according to Definition 1. The sufficiency is hence proved.

The necessity can be shown by reversing the above reasoning, and is left to the reader as an exercise.

2. *Transitivity:* It follows from (3.8) that

$$ss_{VS}(\tilde{A}, \tilde{C}) = \frac{\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{C}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{C}}(x_i))}{\sum_{i=1}^{N} \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{A}}(x_i)} \tag{A.0-3}$$

$$ss_{VS}(\tilde{B}, \tilde{C}) = \frac{\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{B}}(x_i), \overline{\mu}_{\tilde{C}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{B}}(x_i), \underline{\mu}_{\tilde{C}}(x_i))}{\sum_{i=1}^{N} \overline{\mu}_{\tilde{B}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{B}}(x_i)} \tag{A.0-4}$$

When $\tilde{C} \leqslant \tilde{A} \leqslant \tilde{B}$, it follows from Definition 1 that $\underline{\mu}_{\tilde{C}}(x_i) \leqslant \underline{\mu}_{\tilde{A}}(x_i) \leqslant \underline{\mu}_{\tilde{B}}(x_i)$ and $\overline{\mu}_{\tilde{C}}(x_i) \leqslant \overline{\mu}_{\tilde{A}}(x_i) \leqslant \overline{\mu}_{\tilde{B}}(x_i)$ for $\forall x_i$, and hence

$$\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{C}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{C}}(x_i)) = \sum_{i=1}^{N} \overline{\mu}_{\tilde{C}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{C}}(x_i) \tag{A.0-5}$$

$$\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{B}}(x_i), \overline{\mu}_{\tilde{C}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{B}}(x_i), \underline{\mu}_{\tilde{C}}(x_i)) = \sum_{i=1}^{N} \overline{\mu}_{\tilde{C}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{C}}(x_i) \tag{A.0-6}$$

$$\sum_{i=1}^{N} \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{A}}(x_i) \leqslant \sum_{i=1}^{N} \overline{\mu}_{\tilde{B}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{B}}(x_i) \tag{A.0-7}$$

Consequently, $ss_{VS}(\tilde{A}, \tilde{C}) \geqslant ss_{VS}(\tilde{B}, \tilde{C})$ for any $\tilde{C}$.

Similarly, it follows from (3.8) that

$$ss_{VS}(\tilde{C}, \tilde{A}) = \frac{\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{C}}(x_i), \overline{\mu}_{\tilde{A}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{C}}(x_i), \underline{\mu}_{\tilde{A}}(x_i))}{\sum_{i=1}^{N} \overline{\mu}_{\tilde{C}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{C}}(x_i)} \tag{A.0-8}$$

$$ss_{VS}(\tilde{C}, \tilde{B}) = \frac{\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{C}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{C}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{N} \overline{\mu}_{\tilde{C}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{C}}(x_i)} \tag{A.0-9}$$

When $\tilde{A} \leqslant \tilde{B}$, it follows from Definition 1 that $\underline{\mu}_{\tilde{A}}(x_i) \leqslant \underline{\mu}_{\tilde{B}}(x_i)$ and $\overline{\mu}_{\tilde{A}}(x_i) \leqslant \overline{\mu}_{\tilde{B}}(x_i)$ for $\forall x_i$; hence,

$$\begin{aligned} &\min(\overline{\mu}_{\tilde{C}}(x_i), \overline{\mu}_{\tilde{A}}(x_i)) \leqslant \min(\overline{\mu}_{\tilde{C}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) \\ &\min(\underline{\mu}_{\tilde{C}}(x_i), \underline{\mu}_{\tilde{A}}(x_i)) \leqslant \min(\underline{\mu}_{\tilde{C}}(x_i), \underline{\mu}_{\tilde{B}}(x_i)) \end{aligned} \tag{A.0-10}$$

so that

$$\begin{aligned} &\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{C}}(x_i), \overline{\mu}_{\tilde{A}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{C}}(x_i), \underline{\mu}_{\tilde{A}}(x_i)) \\ \leqslant &\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{C}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{C}}(x_i), \underline{\mu}_{\tilde{B}}(x_i)) \end{aligned} \tag{A.0-11}$$

Consequently, $ss_{VS}(\tilde{C}, \tilde{A}) \leqslant ss_{VS}(\tilde{C}, \tilde{B})$ for any $\tilde{C}$.

3. *Overlapping:* According to Definition 2, $\tilde{A}$ and $\tilde{B}$ overlap if $\min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) > 0$ for at least one x_i; hence, if $\tilde{A} \cap \tilde{B} \neq \emptyset$, then

$$\begin{aligned} ss_{VS}(\tilde{A}, \tilde{B}) &= \frac{\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{N} \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{A}}(x_i)} \\ &\geqslant \frac{\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i))}{\sum_{i=1}^{N} \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=1}^{N} \underline{\mu}_{\tilde{A}}(x_i)} \geqslant 0 \end{aligned} \tag{A.0-12}$$

On the other hand, according to Definition 2, when $\tilde{A} \cap \tilde{B} = \emptyset$, $\sum_{i=1}^{N} \min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) = 0$ and $\sum_{i=1}^{N} \min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i)) = 0$; consequently, $ss_{VS}(\tilde{A}, \tilde{B}) = 0$.

Bibliography

[1] K. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets and Systems 20 (1986) 87–97.

[2] K. Atanassov, G. Gargov, Interval valued intuitionistic fuzzy sets, Fuzzy Sets and Systems 31 (1989) 343–349.

[3] C. Cornelis, E. Kerre, Inclusion measures in intuitionistic fuzzy set theory, Lecture Notes in Computer Science 2711 (2004) 345–356.

[4] B. V. Dasarathy, Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques, IEEE Computer Society Press, Los Alamitos, CA, 1991.

[5] A. De Luca, S. Termini, A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, Information and Control 20 (1972) 301–312.

[6] J. Fan, W. Xie, J. Pei, Subsethood measure: New definitions, Fuzzy Sets and Systems 106 (1999) 210–209.

[7] F. Herrera, L. Martinez, A 2-tuple fuzzy linguistic representation model for computing with words, IEEE Trans. on Fuzzy Systems 8 (6) (2000) 746–752.

[8] P. Jaccard, Nouvelles recherches sur la distribution florale, Bulletin de la Societe de Vaud des Sciences Naturelles 44 (1908) 223.

[9] J. Kacprzyk, S. Zadrożny, Computing with words in intelligent database querying: Standalone and internet-based applications, Information Sciences 34 (2001) 71–109.

[10] A. Kaufmann, Introduction to the Theory of Fuzzy Subsets: Fundamental Theoretical Elements, Academic Press, NY, 1975.

[11] G. J. Klir, T. A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice Hall, Englewood Cliffs, NJ, 1988.

[12] B. Kosko, Fuzziness vs. probability, Int'l J. General Systems 17 (1990) 211–240.

[13] J. Lawry, A methodology for computing with words, International Journal of Approximate Reasoning 28 (2001) 51–89.

[14] F. Liu, J. M. Mendel, Aggregation using the fuzzy weighted average, as computed using the Karnik-Mendel Algorithms, IEEE Trans. on Fuzzy Systems 12 (1) (2008) 1–12.

[15] F. Liu, J. M. Mendel, Encoding words into interval type-2 fuzzy sets using an interval approach, IEEE Trans. on Fuzzy Systems 16 (6) (2008) 1503–1521.

[16] X. Liu, S. Zheng, F. Xiong, Entropy and subsethood for general interval-valued intuitionistic fuzzy sets, Lecture Notes in Computer Science 3613 (2005) 42–52.

[17] Y.-H. Liu, F.-L. Xiong, Subsethood on intuitionistic fuzzy sets, in: Proc. 1st Int'l Conf. on Machine Learning and Cybernetics, vol. 3, Beijing, China, 2002.

[18] M. Margaliot, G. Langholz, Fuzzy control of a benchmark problem: A computing with words approach, IEEE Trans. on Fuzzy Systems 12 (2) (2004) 230–235.

[19] O. Martin, G. J. Klir, On the problem of retranslation in computing with perceptions, International Journal of General Systems 35 (6) (2006) 655–674.

[20] D. McKenzie, R. Forsyth, Classification by similarity: An overview of statistical methods of case-based reasoning, Computers in Human Behavior 11 (2) (1995) 273–288.

[21] J. M. Mendel, Computing with words, when words can mean different things to different people, in: Proc. 3rd Int'l ICSC Symposium on Fuzzy Logic and Applications, Rochester, NY, 1999.

[22] J. M. Mendel, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice-Hall, Upper Saddle River, NJ, 2001.

[23] J. M. Mendel, An architecture for making judgments using computing with words, Int'l Journal of Applied Mathematics and Computer Science 12 (3) (2002) 325–335.

[24] J. M. Mendel, Computing with words and its relationships with fuzzistics, Information Sciences 177 (2007) 988–1006.

[25] J. M. Mendel, Computing with words: Zadeh, Turing, Popper and Occam, IEEE Computational Intelligence Magazine 2 (2007) 10–17.

[26] 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.

[27] J. M. Mendel, R. I. John, Type-2 fuzzy sets made simple, IEEE Trans. on Fuzzy Systems 10 (2) (2002) 117–127.

[28] J. M. Mendel, R. I. John, F. Liu, Interval type-2 fuzzy logic systems made simple, IEEE Trans. on Fuzzy Systems 14 (6) (2006) 808–821.

[29] J. M. Mendel, D. Wu, Perceptual reasoning for perceptual computing, IEEE Trans. on Fuzzy Systems 16 (6) (2008) 1550–1564.

[30] J. M. Mendel, D. Wu, Computing with words for hierarchical and distributed decision making, in: D. Ruan (ed.), Computational Intelligence in Complex Decision Systems, Atlantis Press, Paris, France, 2009.

[31] J. M. Mendel, D. Wu, Perceptual Computing: Aiding People in Making Subjective Judgments, Wiley-IEEE Press, Hoboken, NJ, June 2010.

[32] H. T. Nguyen, V. Kreinovich, Computing degrees of subsethood and similarity for interval-valued fuzzy sets: Fast algorithms, in: Proc. 9th Int'l Conf. on Intelligent Technologies, Samui, Thailand, 2008.

[33] R. Nosofsky, Choice, similarity, and the context theory of classification, Journal of experimental psychology. Learning, memory, and cognition 10 (1) (1984) 104–114.

[34] S. K. Pal, L. Polkowski, A. Skowron (eds.), Rough-neural Computing: Techniques for Computing with Words, Springer-Verlag, Heidelberg, Germany, 2003.

[35] J. T. Rickard, J. Aisbett, G. Gibbon, Fuzzy subsethood for fuzzy sets of type-2 and generalized type-n, IEEE Trans. on Fuzzy Systems 17 (1) (2009) 50–60.

[36] J. T. Rickard, J. Aisbett, G. Gibbon, D. Morgenthaler, Fuzzy subsethood for type-n fuzzy sets, in: Proc. NAFIPS, New York, 2008.

[37] S. H. Rubin, Computing with words, IEEE Trans. on Systems, Man and Cybernetics–B 29 (4) (1999) 518 – 524.

[38] D. Sinha, E. R. Dougherty, Fuzzification of set inclusion: Theory and applications, Fuzzy Sets and Systems 55 (1993) 15–42.

[39] I. B. Türkşen, Interval-valued fuzzy sets based on normal forms, Fuzzy Sets and Systems 20 (1986) 191–210.

[40] I. B. Türkşen, Type 2 representation and reasoning for CWW, Fuzzy Sets and Systems 127 (2002) 17–36.

[41] I. Vlachos, G. Sergiadis, Subsethood, entropy, and cardinality for interval-valued fuzzy sets – an algebraic derivation, Fuzzy Sets and Systems 158 (2007) 1384–1396.

[42] T. S. Wallsten, D. V. Budescu, A review of human linguistic probability processing: General principles and empirical evidence, The Knowledge Engineering Review 10 (1) (1995) 43–62.

[43] H. Wang, D. Qiu, Computing with words via Turing machines: A formal approach, IEEE Trans. on Fuzzy Systems 11 (6) (2003) 742–753.

[44] J.-H. Wang, J. Hao, A new version of 2-tuple fuzzy linguistic representation model for computing with words, IEEE Trans. on Fuzzy Systems 14 (3) (2006) 435–445.

[45] P. Wang (ed.), Computing With Words, John Wiley & Sons, New York, 2001.

[46] D. Wu, Intelligent systems for decision support, Ph.D. thesis, University of Southern California, Los Angeles, CA (May 2009).

[47] D. Wu, J. M. Mendel, The linguistic weighted average, in: Proc. FUZZ-IEEE, Vancouver, BC, Canada, 2006.

[48] D. Wu, J. M. Mendel, Aggregation using the linguistic weighted average and interval type-2 fuzzy sets, IEEE Trans. on Fuzzy Systems 15 (6) (2007) 1145–1161.

[49] D. Wu, J. M. Mendel, Uncertainty measures for interval type-2 fuzzy sets, Information Sciences 177 (23) (2007) 5378–5393.

[50] D. Wu, J. M. Mendel, Corrections to 'Aggregation using the linguistic weighted average and interval type-2 fuzzy sets', IEEE Trans. on Fuzzy Systems 16 (6) (2008) 1664–1666.

[51] D. Wu, J. M. Mendel, A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets, Information Sciences 179 (8) (2009) 1169–1192.

[52] D. Wu, J. M. Mendel, Computing with words for hierarchical decision making applied to evaluating a weapon system, *submitted for publication in* IEEE Trans. on Fuzzy Systems.

[53] D. Wu, J. M. Mendel, Perceptual reasoning for perceptual computing: A similarity-based approach, IEEE Trans. on Fuzzy Systems 17 (6) (2009) 1397–1411.

[54] D. Wu, J. M. Mendel, Similarity-based perceptual reasoning for perceptual computing, in: Proc. FUZZ-IEEE, Jeju Island, South Korea, 2009.

[55] R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Trans. on Systems, Man and Cybernetics 18 (1988) 183–190.

[56] R. Yager, On the retranslation process in Zadeh's paradigm of computing with words, IEEE Trans. on Systems, Man and Cybernetics–B 34 (2) (2004) 1184–1195.

[57] R. Yager, J. Kacprzyk, The Ordered Weighted Averaging Operators: Theory and Applications, Kluwer, Norwell, MA, 1997.

[58] R. R. Yager, On the measure of fuzziness and negation part I: Membership in the unit interval, International Journal of General Systems 5 (1979) 221–229.

[59] R. R. Yager, On the measure of fuzziness and negation part II: Lattices, Information and Control 44 (1980) 236–260.

[60] R. R. Yager, Entropy and specificity in a mathematical theory of evidence, International Journal of General Systems 9 (1983) 249–260.

[61] M.-S. Yang, D.-C. Lin, On similarity and inclusion measures between type-2 fuzzy sets with an application to clustering, Computers and Mathematics with Applications 57 (2009) 896–907.

[62] M. Ying, A formal model of computing with words, IEEE Trans. on Fuzzy Systems 10 (5) (2002) 640–652.

[63] V. R. Young, Fuzzy subsethood, Fuzzy Sets and Systems 77 (1996) 371–384.

[64] L. A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.

[65] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning-1, Information Sciences 8 (1975) 199–249.

[66] L. A. Zadeh, Fuzzy logic = Computing with words, IEEE Trans. on Fuzzy Systems 4 (1996) 103–111.

[67] L. A. Zadeh, From computing with numbers to computing with words – From manipulation of measurements to manipulation of perceptions, IEEE Trans. on Circuits and Systems–I: Fundamental Theory and Applications 46 (1) (1999) 105–119.

[68] L. A. Zadeh, Toward a generalized theory of uncertainty (GTU) – An outline, Information Sciences 172 (2005) 1–40.

[69] L. A. Zadeh, Toward human level machine intelligence – Is it achievable? The need for a paradigm shift, IEEE Computational Intelligence Magazine 3 (3) (2008) 11–22.

[70] L. A. Zadeh, J. Kacprzyk (eds.), Computing with words in information/intelligent systems: 1. Foundations, 2. Applications, Physica-Verlag, Heidelberg, 1999.

[71] W. Zeng, H. Li, Inclusion measures, similarity measures, and the fuzziness of fuzzy sets and their relations, International Journal of Intelligent Systems 21 (2006) 639–653.