Information Sciences 179 (2009) 1169–1192

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 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

ARTICLE INFO

Article history:
Received 7 February 2008
Received in revised form 20 November 2008
Accepted 14 December 2008

Keywords:
Interval type-2 fuzzy sets
Ranking methods
Similarity measures
Uncertainty measures
Computing with words

ABSTRACT

Ranking methods, similarity measures and uncertainty measures are very important concepts for interval type-2 fuzzy sets (IT2 FSs). So far, there is only one ranking method for such sets, whereas there are many similarity and uncertainty measures. A new ranking method and a new similarity measure for IT2 FSs are proposed in this paper. All these ranking methods, similarity measures and uncertainty measures are compared based on real survey data and then the most suitable ranking method, similarity measure and uncertainty measure that can be used in the computing with words paradigm are suggested. The results are useful in understanding the uncertainties associated with linguistic terms and hence how to use them effectively in survey design and linguistic information processing.

1. Introduction

Zadeh coined the phrase "*computing with words*" (CWW) [43,44]. According to [44], CWW is "*a methodology in which the objects of computation are words and propositions drawn from a natural language.*" There are at least two types of uncertainties associated with a word [29]: intra-personal uncertainty and inter-personal uncertainty. The former is explicitly pointed out by Wallsten and Budescu [29] 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 modeling it by type-1 fuzzy sets (T1 FSs). The latter is pointed out by Mendel [13] as "*words mean different things to different people*" and Wallsten and Budescu [29] as "*different individuals use diverse expressions to describe identical situations and understand the same phrases differently when hearing or reading them.*" Because each interval type-2 FS (IT2 FS) can be viewed as a group of T1 FSs and hence can model both types of uncertainty, we suggest using IT2 FSs in CWW [14,13,17].

CWW using T1 FSs have been studied by many authors, including Tong and Bonissone [28], Schmucker [26], Zadeh [43], Buckley and Feuring [2], Yager [38,41], Margaliot and Langholz [12], Novak [25], etc., though some of them did not call it CWW. Mendel was the first to study CWW using IT2 FSs [15,16], and he proposed [16] a specific architecture (Fig. 1) for making judgments by CWW. It is called a *perceptual computer*—Per-C for short. 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 recommendation, which can be in the form of word, rank, or class. When a word recommendation is desired, usually a vocabulary (codebook)

* Corresponding author. Tel.: +1 213 740 4456.
E-mail addresses: dongruiw@usc.edu (D. Wu), mendel@sipi.usc.edu (J.M. Mendel).

[1] Zadeh calls this *constraint explicitation* in [43,44]. In some of his recent talks, he calls this *precisiation*.
[2] Zadeh calls this *linguistic approximation* in [43,44].

0020-0255/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2008.12.010

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.

To operate the Per-C, we need to solve the following problems:

(i) How to transform linguistic perceptions into IT2 FSs, i.e. the encoding problem. Two approaches have appeared in the literature: the *person membership function (MF) approach* [17] and the *interval end-points approach* [20,22]. Recently, Liu and Mendel [11] proposed a new method called the *interval approach*, which captures the strong points of both the person-MF and interval end-points approaches.
(ii) How to construct the CWW engine, which maps IT2 FSs into IT2 FSs. There may be different kinds of CWW engines, e.g., the linguistic weighted average[3] (LWA) [32,33,35], perceptual reasoning (PR) [18,19], etc.
(iii) How to map the output of the CWW engine into a word recommendation (linguistic label). To map an IT2 FS into a word, it must be possible to compare the similarity between two IT2 FSs. There are five existing similarity measures for IT2 FSs in the literature [3,5,23,37,45].
(iv) How to rank the outputs of the CWW engine. Ranking is needed when several alternatives are compared to find the best. Because the performance of each alternative is represented by an IT2 FS obtained from the CWW engine, a ranking method for IT2 FSs is needed. Only one such method has been proposed so far by Mitchell [24].
(v) How to quantify the uncertainty associated with an IT2 FS. As pointed out by Klir [9], "*once uncertainty (and information) measures become well justified, they can very effectively be utilized for managing uncertainty and the associated information. For example, they can be utilized for extrapolating evidence, assessing the strength of relationship between given groups of variables, assessing the influence of given input variables on given output variables, measuring the loss of information when a system is simplified, and the like.*" Several basic principles of uncertainty have been proposed [6,9], e.g., the principles of minimum uncertainty, maximum uncertainty, and uncertainty invariance. Five uncertainty measures have been proposed in [34]; however, an open problem is which one to use.

Only problems (iii)–(v) are considered in this paper. Our objectives are to:

(i) Evaluate ranking methods, similarity measures and uncertainty measures for IT2 FSs based on real survey data; and,
(ii) Suggest the most suitable ranking method, similarity measure and uncertainty measure that can be used in the Per-C instantiation of the CWW paradigm.

The rest of this paper is organized as follows: Section 2 presents the 32 word FOUs used in this study. Section 3 proposes a new ranking method for IT2 FSs and compares it with Mitchell's method. Section 4 proposes a new similarity measure for IT2 FSs and compares it with the existing five methods. Section 5 computes uncertainty measures for the 32 words and studies their relationships. Section 6 draws conclusions.

2. Word FOUs

The dataset used herein was collected from 28 subjects at the Jet Propulsion Laboratory[4] (JPL). Thirty-two words were randomly ordered and presented to the subjects. Each subject was asked to provide the end-points of an interval for each word on the scale 0–10. The 32 words can be grouped into three classes: small-sounding words (*little, low amount, somewhat small, a smidgen, none to very little, very small, very little, teeny-weeny, small amount* and *tiny*), medium-sounding words (*fair amount, modest amount, moderate amount, medium, good amount, a bit, some to moderate* and *some*), and large-sounding words (*sizeable, large, quite a bit, humongous amount, very large, extreme amount, considerable amount, a lot, very sizeable, high amount, maximum amount, very high amount* and *substantial amount*).

Liu and Mendel's interval approach for word modeling [11] was used to map these data intervals into footprints of uncertainty (FOUs). For each word, after some pre-processing, during which some intervals (e.g., outliers) were eliminated, each of the remaining intervals was classified as either an interior, left-shoulder or right-shoulder IT2 FS. Then, each of the word's data intervals was individually mapped into its respective T1 interior, left-shoulder or right-shoulder MF, after which the union of all of these T1 MFs was taken, and the union was upper and lower bounded. The result is an FOU for an IT2 FS model of the word, which is completely described by these lower and upper bounds, called the lower membership function (LMF) and the upper membership function (UMF), respectively. The 32 word FOUs are depicted in Fig. 2, and their parameters are shown in Table 1. The actual survey data for the 32 words and the software are available online at http://sipi.usc.edu/~mendel/software.

Note that although all of our numerical computations and results are for the Fig. 2 FOUs and Table 1 data, they can easily be re-computed for new data. Note also that the 32 word vocabulary can be partitioned into several smaller sub-vocabularies, each of which covers the domain [0,10]. Some examples of the sub-vocabularies are given in [11]. All of our numerical computations can be repeated for these sub-vocabularies.

[3] An LWA is expressed as $\widetilde{Y} = \sum_{i=1}^{N} \widetilde{X}_i \widetilde{W}_i / \sum_{i=1}^{N} \widetilde{W}_i$ where $\widetilde{X}_i$ and $\widetilde{W}_i$ are words modeled by IT2 FSs.
[4] This was done in 2002 when Mendel gave an in-house short course on fuzzy sets and systems at JPL.

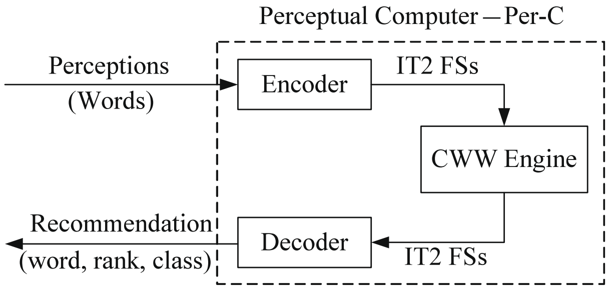

Fig. 1. Conceptual structure of CWW.

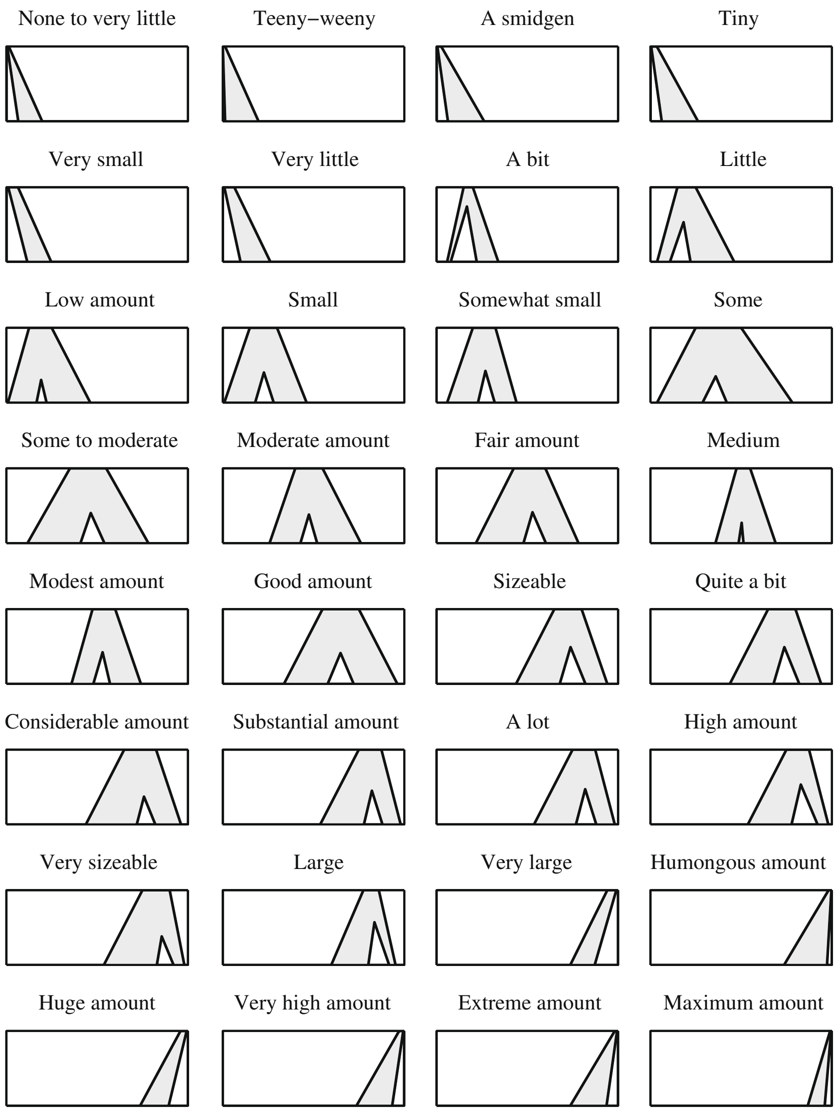

Fig. 2. The 32 word FOUs ranked by their centers of centroid. To read this figure, scan from left to right starting at the top of the page.

3. Ranking methods for IT2 FSs

Though there are more than 35 reported different methods for ranking type-1 fuzzy numbers [30,31], to the best knowledge of the authors, only one method on ranking IT2 FSs has been published, namely Mitchell's method in [24]. We will first

Table 1
Parameters of the 32 word FOUs. As shown in Fig. 3, each UMF is represented by (a, b, c, d), and each LMF is represented (e, f, g, i, h).

Word	UMF	LMF	$C(\widetilde{A}_i)$	$c(\widetilde{A}_i)$
1. None to very little	[0, 0, 0.14, 1.97]	[0, 0, 0.05, 0.66, 1]	[0.22, 0.73]	0.47
2. Teeny-weeny	[0, 0, 0.14, 1.97]	[0, 0, 0.01, 0.13, 1]	[0.05, 1.07]	0.56
3. A smidgen	[0, 0, 0.26, 2.63]	[0, 0, 0.05, 0.63, 1]	[0.21, 1.05]	0.63
4. Tiny	[0, 0, 0.36, 2.63]	[0, 0, 0.05, 0.63, 1]	[0.21, 1.06]	0.64
5. Very small	[0, 0, 0.64, 2.47]	[0, 0, 0.10, 1.16, 1]	[0.39, 0.93]	0.66
6. Very little	[0, 0, 0.64, 2.63]	[0, 0, 0.09, 0.99, 1]	[0.33, 1.01]	0.67
7. A bit	[0.59, 1.50, 2.00, 3.41]	[0.79, 1.68, 1.68, 2.21, 0.74]	[1.42, 2.08]	1.75
8. Little	[0.38, 1.50, 2.50, 4.62]	[1.09, 1.83, 1.83, 2.21, 0.53]	[1.31, 2.95]	2.13
9. Low amount	[0.09, 1.25, 2.50, 4.62]	[1.67, 1.92, 1.92, 2.21, 0.30]	[0.92, 3.46]	2.19
10. Small	[0.09, 1.50, 3.00, 4.62]	[1.79, 2.28, 2.28, 2.81, 0.40]	[1.29, 3.34]	2.32
11. Somewhat small	[0.59, 2.00, 3.25, 4.41]	[2.29, 2.70, 2.70, 3.21, 0.42]	[1.76, 3.43]	2.59
12. Some	[0.38, 2.50, 5.00, 7.83]	[2.88, 3.61, 3.61, 4.21, 0.35]	[2.04, 5.77]	3.90
13. Some to moderate	[1.17, 3.50, 5.50, 7.83]	[4.09, 4.65, 4.65, 5.41, 0.40]	[3.02, 6.11]	4.56
14. Moderate amount	[2.59, 4.00, 5.50, 7.62]	[4.29, 4.75, 4.75, 5.21, 0.38]	[3.74, 6.16]	4.95
15. Fair amount	[2.17, 4.25, 6.00, 7.83]	[4.79, 5.29, 5.29, 6.02, 0.41]	[3.85, 6.41]	5.13
16. Medium	[3.59, 4.75, 5.50, 6.91]	[4.86, 5.03, 5.03, 5.14, 0.27]	[4.19, 6.19]	5.19
17. Modest amount	[3.59, 4.75, 6.00, 7.41]	[4.79, 5.30, 5.30, 5.71, 0.42]	[4.57, 6.24]	5.41
18. Good amount	[3.38, 5.50, 7.50, 9.62]	[5.79, 6.50, 6.50, 7.21, 0.41]	[5.11, 7.89]	6.50
19. Sizeable	[4.38, 6.50, 8.00, 9.41]	[6.79, 7.38, 7.38, 8.21, 0.49]	[6.17, 8.15]	7.16
20. Quite a bit	[4.38, 6.50, 8.00, 9.41]	[6.79, 7.38, 7.38, 8.21, 0.49]	[6.17, 8.15]	7.16
21. Considerable amount	[4.38, 6.50, 8.25, 9.62]	[7.19, 7.58, 7.58, 8.21, 0.37]	[5.97, 8.52]	7.25
22. Substantial amount	[5.38, 7.50, 8.75, 9.81]	[7.79, 8.22, 8.22, 8.81, 0.45]	[6.95, 8.86]	7.90
23. A lot	[5.38, 7.50, 8.75, 9.83]	[7.69, 8.19, 8.19, 8.81, 0.47]	[6.99, 8.83]	7.91
24. High amount	[5.38, 7.50, 8.75, 9.81]	[7.79, 8.30, 8.30, 9.21, 0.53]	[7.19, 8.82]	8.01
25. Very sizeable	[5.38, 7.50, 9.00, 9.81]	[8.29, 8.56, 8.56, 9.21, 0.38]	[6.95, 9.10]	8.03
26. Large	[5.98, 7.75, 8.60, 9.52]	[8.03, 8.36, 8.36, 9.17, 0.57]	[7.50, 8.75]	8.12
27. Very large	[7.37, 9.41, 10, 10]	[8.72, 9.91, 10, 10, 1]	[9.03, 9.57]	9.30
28. Humongous amount	[7.37, 9.82, 10, 10]	[9.74, 9.98, 10, 10, 1]	[8.70, 9.91]	9.31
29. Huge amount	[7.37, 9.59, 10, 10]	[8.95, 9.93, 10, 10, 1]	[9.03, 9.65]	9.34
30. Very high amount	[7.37, 9.73, 10, 10]	[9.34, 9.95, 10, 10, 1]	[8.96, 9.78]	9.37
31. Extreme amount	[7.37, 9.82, 10, 10]	[9.37, 9.95, 10, 10, 1]	[8.96, 9.79]	9.38
32. Maximum amount	[8.68, 9.91, 10, 10]	[9.61, 9.97, 10, 10, 1]	[9.50, 9.87]	9.69

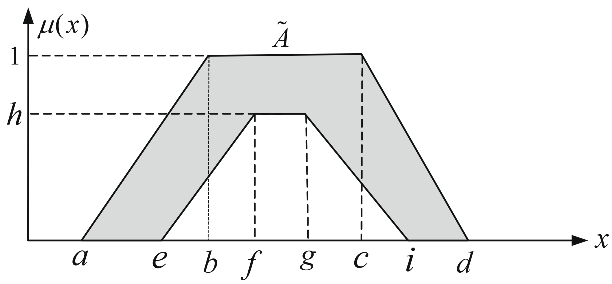

Fig. 3. The nine points to determine an FOU. (a, b, c, d) determines a normal trapezoidal UMF, and (e, f, g, i, h) determines a trapezoidal LMF with height h.

introduce some reasonable ordering properties for IT2 FSs, and then compare Mitchell's method against them. A new ranking method for IT2 FSs is proposed at the end of this section.

3.1. Reasonable ordering properties for IT2 FSs

Wang and Kerre [30,31] performed a comprehensive study of T1 FSs ranking methods based on seven reasonable ordering properties for T1 FSs. When extended to IT2 FSs, these properties are[5]:

[P1.] If $\widetilde{A} \succeq \widetilde{B}$ and $\widetilde{B} \succeq \widetilde{A}$, then $\widetilde{A} \sim \widetilde{B}$.
[P2.] If $\widetilde{A} \succeq \widetilde{B}$ and $\widetilde{B} \succeq \widetilde{C}$, then $\widetilde{A} \succeq \widetilde{C}$.
[P3.] If $\widetilde{A} \cap \widetilde{B} = \emptyset$ and $\widetilde{A}$ is on the right of $\widetilde{B}$, then $\widetilde{A} \succeq \widetilde{B}$.
[P4.] The order of $\widetilde{A}$ and $\widetilde{B}$ is not affected by the other IT2 FSs under comparison.

[5] There is another property saying that "for any IT2 FS $\widetilde{A}$, $\widetilde{A} \succeq \widetilde{A}$;" however, it is not included here since it sounds weird, though our centroid-based ranking method satisfies it.

[P5.] If $\widetilde{A} \succeq \widetilde{B}$, then[6] $\widetilde{A} + \widetilde{C} \succeq \widetilde{B} + \widetilde{C}$.
[P6.] If $\widetilde{A} \succeq \widetilde{B}$, then[7] $\widetilde{A}\widetilde{C} \succeq \widetilde{B}\widetilde{C}$.

where $\succeq$ means "larger than or equal to in the sense of ranking," $\sim$ means "the same rank," and $\widetilde{A} \cap \widetilde{B} = \emptyset$ is defined in:

Definition 1. $\widetilde{A} \cap \widetilde{B} \neq \emptyset$, i.e., $\widetilde{A}$ and $\widetilde{B}$ overlap, if $\exists x$ such that $\min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) > 0$. $\widetilde{A} \cap \widetilde{B} = \emptyset$, i.e., $\widetilde{A}$ and $\widetilde{B}$ do not overlap, if $\min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) = 0$ for $\forall x$.

All the six properties are intuitive. P4 may look trivial, but it is worth emphasizing because some ranking methods [30,31] first set up reference set(s) and then all FSs are compared with the reference set(s). The reference set(s) may depend on the FSs under consideration, so it is possible (but not desirable) that $\widetilde{A} \succeq \widetilde{B}$ when $\{\widetilde{A}, \widetilde{B}, \widetilde{C}\}$ are ranked whereas $\widetilde{A} \prec \widetilde{B}$ when $\{\widetilde{A}, \widetilde{B}, \widetilde{D}\}$ are ranked.

3.2. Mitchell's method for ranking IT2 FSs

Mitchell [24] proposed a ranking method for general type-2 FSs. When specialized to M IT2 FSs $\widetilde{A}_m$ ($m = 1, \ldots, M$), the procedure is:

(i) Discretize the primary variable's universe of discourse, X, into N points, that are used by all $\widetilde{A}_m$, $m = 1, \ldots, M$.
(ii) Find H random embedded T1 FSs[8], A_e^{mh}, $h = 1, \ldots, H$, for each of the M IT2 FSs $\widetilde{A}_m$, as:

$$\mu_{A_e^{mh}}(x_n) = r_{mh}(x_n) \times [\bar{\mu}_{\widetilde{A}_m}(x_n) - \underline{\mu}_{\widetilde{A}_m}(x_n)] + \underline{\mu}_{\widetilde{A}_m}(x_n) \quad n = 1, 2, \ldots, N \tag{1}$$

where $r_{mh}(x_n)$ is a random number chosen uniformly in $[0, 1]$, and $\underline{\mu}_{\widetilde{A}_m}(x_n)$ and $\bar{\mu}_{\widetilde{A}_m}(x_n)$ are the lower and upper memberships of $\widetilde{A}_m$ at x_n.
(iii) Form the H^M different combinations of $\{A_e^{1h}, A_e^{2h}, \ldots, A_e^{Mh}\}_i$, $i = 1, \ldots, H^M$.
(iv) Use a T1 FS ranking method to rank each of the M^H $\{A_e^{1h}, A_e^{2h}, \ldots, A_e^{Mh}\}_i$. Denote the rank of A_{mh}^e in $\{A_e^{1h}, A_e^{2h}, \ldots, A_e^{Mh}\}_i$ as r_{mi}.
(v) Compute the final rank of $\widetilde{A}_m$ as

$$r_m = \frac{1}{H^M} \sum_{i=1}^{H^M} r_{mi}, \quad m = 1, \ldots, M \tag{2}$$

Observe from the above procedure that:

(i) The output ranking, r_m, is a crisp number; however, usually it is not an integer. These r_m ($m = 1, \ldots, M$) need to be sorted in order to find the correct ranking.
(ii) A total of H^M T1 FS rankings must be evaluated before r_m can be computed. For our problem, where 32 IT2 FSs have to be ranked, even if H is chosen as a small number, say 2, $2^{32} \approx 4.295 \times 10^9$ T1 FS rankings have to be evaluated, and each evaluation involves 32 T1 FSs. This is highly impractical. Although two fast algorithms are proposed in [24], because our FOUs have lots of overlap, the computational cost cannot be reduced significantly. Note also that choosing the number of realizations H as 2 is not meaningful; it should be much larger, and for larger H, the number of rankings becomes astronomical.
(iii) Because there are random numbers involved, r_m is random and will change from experiment to experiment. When H is large, some kind of stochastic convergence can be expected to occur for r_m (e.g., convergence in probability); however, as mentioned in (ii), the computational cost is prohibitive.
(iv) Because of the random nature of Mitchell's ranking method, it only satisfies P3 of the six reasonable properties proposed in Section 3.1.

3.3. A new centroid-based ranking method

A simple ranking method based on the centroids of IT2 FSs is proposed in this subsection.

---

[6] $\widetilde{A} + \widetilde{C}$ is computed using α-cuts [10] and Extension Principle [42], i.e., let $\widetilde{A}^\alpha$, $\widetilde{C}^\alpha$ and $(\widetilde{A} + \widetilde{C})^\alpha$ be α-cuts on $\widetilde{A}$, $\widetilde{C}$ and $\widetilde{A} + \widetilde{C}$, respectively; then, $(\widetilde{A} + \widetilde{C})^\alpha = \widetilde{A}^\alpha + \widetilde{C}^\alpha$ for $\forall \alpha \in [0, 1]$.

[7] $\widetilde{A}\widetilde{C}$ is computed using α-cuts [10] and extension principle [42], i.e., let $\widetilde{A}^\alpha$, $\widetilde{C}^\alpha$ and $(\widetilde{A}\widetilde{C})^\alpha$ be α-cuts on $\widetilde{A}$, $\widetilde{C}$ and $\widetilde{A}\widetilde{C}$, respectively; then, $(\widetilde{A}\widetilde{C})^\alpha = \widetilde{A}^\alpha \widetilde{C}^\alpha$ for $\forall \alpha \in [0, 1]$.

[8] Visually, an embedded T1 FS of an IT2 FS is a T1 FS whose membership function lies within the FOU of the IT2 FS. A more precise mathematical definition can be found in [13].

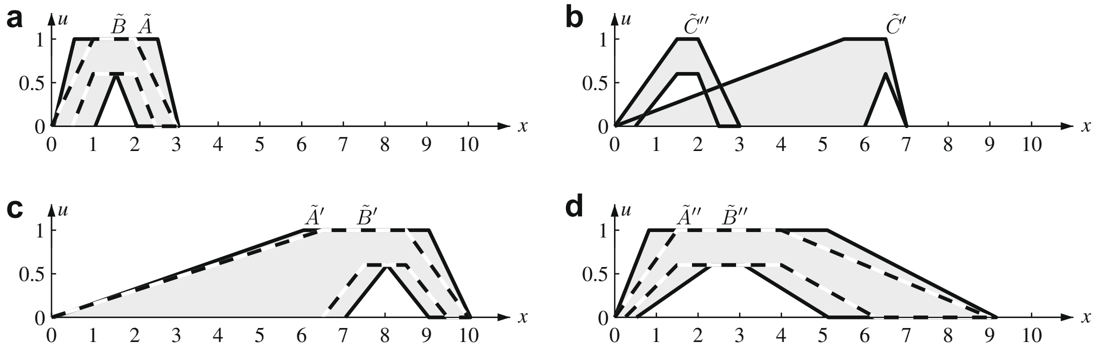

Fig. 4. Counter-examples for P5 and P6. (a) $\tilde{A}$ is the solid curve and $\tilde{B}$ is the dashed curve. $c(\tilde{A}) = 1.55$ and $c(\tilde{B}) = 1.50$ and hence $\tilde{A} \succeq \tilde{B}$. $UMF(\tilde{A}) = [0.05, 0.55, 2.55, 3.05]$, $LMF(\tilde{A}) = [1.05, 1.55, 1.55, 2.05, 0.6]$, $UMF(\tilde{B}) = [0, 1, 2, 3]$ and $LMF(\tilde{B}) = [0.5, 1, 2, 2.5, 0.6]$; (b) $\tilde{C}'$ used in demonstrating P5 and $\tilde{C}''$ used in demonstrating P6. $UMF(\tilde{C}') = [0, 5.5, 6.5, 7]$, $LMF(\tilde{C}') = [6, 6.5, 6.5, 7, 0.6]$, $UMF(\tilde{C}'') = [0, 1.5, 2, 3]$ and $LMF(\tilde{C}'') = [0.5, 1.5, 2, 2.5, 0.6]$; (c) $\tilde{A}' = \tilde{A} + \tilde{C}'$ is the solid curve and $\tilde{B}' = \tilde{B} + \tilde{C}'$ is the dashed curve. $c(\tilde{A}') = 6.53$ and $c(\tilde{B}') = 6.72$ and hence $\tilde{A}' \preceq \tilde{B}'$; (d) $\tilde{A}'' = \tilde{A}\tilde{C}''$ is the solid curve and $\tilde{B}'' = \tilde{B}\tilde{C}''$ is the dashed curve. $c(\tilde{A}'') = 3.44$ and $c(\tilde{B}'') = 3.47$ and hence $\tilde{A}'' \preceq \tilde{B}''$.

Definition 2. [13] The centroid $C(\tilde{A})$ of an IT2 FS $\tilde{A}$ is the union of the centroids of all its embedded T1 FSs A_e, i.e.,

$$C(\tilde{A}) \equiv \bigcup_{\forall A_e} c(A_e) = [c_l(\tilde{A}), c_r(\tilde{A})], \tag{3}$$

where $\bigcup$ is the union operation, and

$$c_l(\tilde{A}) = \min_{\forall A_e} c(A_e) \tag{4}$$

$$c_r(\tilde{A}) = \max_{\forall A_e} c(A_e) \tag{5}$$

$$c(A_e) = \frac{\sum_{i=1}^{N} x_i \mu_{A_e}(x_i)}{\sum_{i=1}^{N} \mu_{A_e}(x_i)}. \tag{6}$$

It has been shown [8,13,21] that $c_l(\tilde{A})$ and $c_r(\tilde{A})$ can be expressed as

$$c_l(\tilde{A}) = \frac{\sum_{i=1}^{L} x_i \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^{N} x_i \underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{L} \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^{N} \underline{\mu}_{\tilde{A}}(x_i)} \tag{7}$$

$$c_r(\tilde{A}) = \frac{\sum_{i=1}^{R} x_i \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^{N} x_i \bar{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{R} \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^{N} \bar{\mu}_{\tilde{A}}(x_i)}. \tag{8}$$

Switch points L and R, as well as $c_l(\tilde{A})$ and $c_r(\tilde{A})$, are computed by iterative KM Algorithms [8,13,36].

Centroid-based ranking method: First compute the average centroid for each IT2 FS $\tilde{A}_i$,

$$c(\tilde{A}_i) = \frac{c_l(\tilde{A}_i) + c_r(\tilde{A}_i)}{2}, \quad i = 1, \ldots, N \tag{9}$$

and then sort $c(\tilde{A}_i)$ to obtain the rank of $\tilde{A}_i$.

The ranking method can be viewed as a generalization of Yager's first ranking method for T1 FSs [39] to IT2 FSs.

Theorem 1. *The centroid-based ranking method satisfies the first four reasonable properties.*

Proof 1. P1–P4 in Section 3.1 are proved in order.

[P1.] $\tilde{A} \succeq \tilde{B}$ means $c(\tilde{A}) \geqslant c(\tilde{B})$ and $\tilde{B} \succeq \tilde{A}$ means $c(\tilde{B}) \geqslant c(\tilde{A})$, and hence $c(\tilde{A}) = c(\tilde{B})$, i.e., $\tilde{A} \sim \tilde{B}$.
[P2.] For the centroid-based ranking method, $\tilde{A} \succeq \tilde{B}$ means $c(\tilde{A}) \geqslant c(\tilde{B})$ and $\tilde{B} \succeq \tilde{C}$ means $c(\tilde{B}) \geqslant c(\tilde{C})$, and hence $c(\tilde{A}) \geqslant c(\tilde{C})$, i.e., $\tilde{A} \succeq \tilde{C}$.
[P3.] If $\tilde{A} \cap \tilde{B} = \emptyset$ and $\tilde{A}$ is on the right of $\tilde{B}$, then $c(\tilde{A}) > c(\tilde{B})$, i.e., $\tilde{A} \succeq \tilde{B}$.
[P4.] Because the order of $\tilde{A}$ and $\tilde{B}$ is completely determined by $c(\tilde{A})$ and $c(\tilde{B})$, which have nothing to do with the other IT2 FSs under comparison, the order of $\tilde{A}$ and $\tilde{B}$ is not affected by the other IT2 FSs. □

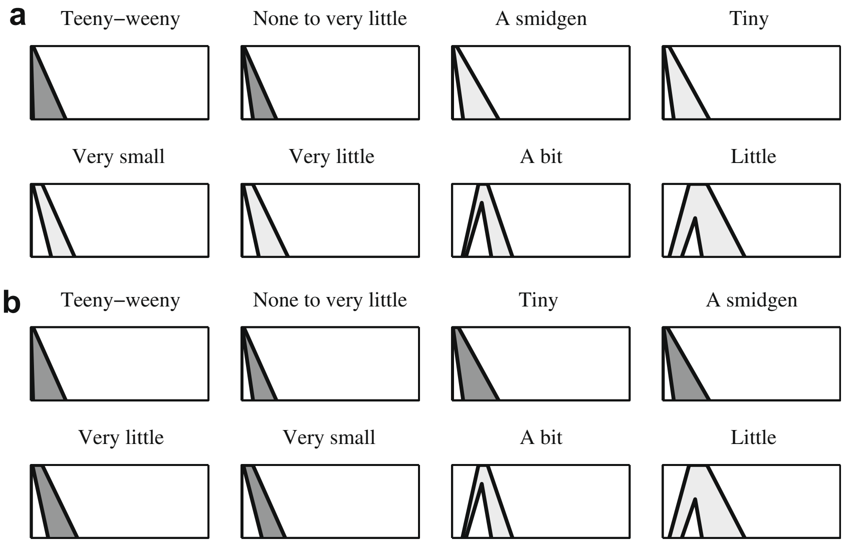

Fig. 5. Ranking of the first eight word FOUs using Mitchell's method: (a) $H = 2$; and (b) $H = 3$.

The centroid-based ranking method does not always satisfy P5 and P6. A counter-example of P5 is shown in Fig. 4, and a counter-example of P6 is shown in Fig. 4; however, they happen only when $c(\widetilde{A})$ and $c(\widetilde{B})$ are very close to each other. For most cases, P5 and P6 are still satisfied. In summary, the centroid-based ranking method satisfies three more of the reasonable ordering properties than Mitchell's method.

3.4. Comparative study

In this section, the performances of the two IT2 FS ranking methods are compared using the 32 word FOUs.

The ranking of the 32 word FOUs using this centroid-based method has already been presented in Fig. 2. Observe that:

(i) The six smallest terms are left-shoulders, the six largest terms are right shoulders, and the terms in-between have interior FOUs.
(ii) Visual examination shows that the ranking is reasonable; it also coincides with the meanings of the words.

Because it is computationally prohibitive to rank all 32 words in one pass using Mitchell's method, only the first eight words in Fig. 2 were used to evaluate Mitchell's method. To be consistent, the T1 FS ranking method used in Mitchell's method is a special case of the centroid-based ranking method for IT2 FSs, i.e., the centroids of the T1 FSs were computed and then were used to rank the corresponding T1 FSs. Ranking results with $H = 2$ and $H = 3$ are shown in Fig. 5a and b, respectively. Words which have a different rank than that in Fig. 2 are shaded more darkly. Observe that:

(i) The ranking is different from that obtained from the centroid-based ranking method.
(ii) The rankings from $H = 2$ and $H = 3$ do not agree.

In summary, the centroid-based ranking method for IT2 FSs seems to be a better choice than Mitchell's method for CWW.

4. Similarity measures

In this section, five existing similarity measures [3,5,23,37,45] for IT2 FSs are briefly reviewed, and then a new similarity measure, having reduced computational cost, is proposed. Before that, a definition is introduced.

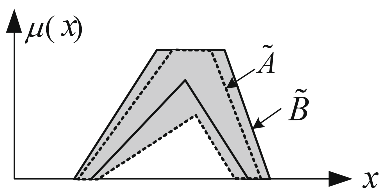

Fig. 6. An illustration of $\widetilde{A} \leqslant \widetilde{B}$.

Definition 3. $\tilde{A} \leqslant \tilde{B}$ if $\bar{\mu}_{\tilde{A}}(x) \leqslant \bar{\mu}_{\tilde{B}}(x)$ and $\underline{\mu}_{\tilde{A}}(x) \leqslant \underline{\mu}_{\tilde{B}}(x)$ for $\forall x \in X$.

An illustration of $\tilde{A} \leqslant \tilde{B}$ is shown in Fig. 6.

The following four properties[9] [37] serve as criteria in the comparisons of the six measures:

[P1.] *Reflexivity:* $s(\tilde{A}, \tilde{B}) = 1 \iff \tilde{A} = \tilde{B}$.
[P2.] *Symmetry:* $s(\tilde{A}, \tilde{B}) = s(\tilde{B}, \tilde{A})$.
[P3.] *Transitivity:* If $\tilde{A} \leqslant \tilde{B} \leqslant \tilde{C}$, then $s(\tilde{A}, \tilde{B}) \geqslant s(\tilde{A}, \tilde{C})$.
[P4.] *Overlapping:* If $\tilde{A} \cap \tilde{B} \neq \emptyset$, then $s(\tilde{A}, \tilde{B}) > 0$; otherwise, $s(\tilde{A}, \tilde{B}) = 0$.

4.1. Mitchell's IT2 FS similarity measure

Mitchell was the first to define a similarity measure for *general* T2 FSs [23]. For the purpose of this paper, only its special case is explained, when both $\tilde{A}$ and $\tilde{B}$ are IT2 FSs:

(i) Discretize the primary variable's universe of discourse, X, into N points, that are used by both $\tilde{A}$ and $\tilde{B}$.
(ii) Find H embedded T1 FSs for IT2 FS $\tilde{A}$ ($h = 1, 2, \ldots, H$), i.e.

$$\mu_{A_e^h}(x_n) = r_h(x_n) \times [\bar{\mu}_{\tilde{A}}(x_n) - \underline{\mu}_{\tilde{A}}(x_n)] + \underline{\mu}_{\tilde{A}}(x_n), \qquad n = 1, 2, \ldots, N \tag{10}$$

where $r_h(x_n)$ is a random number chosen uniformly in $[0, 1]$, and $\underline{\mu}_{\tilde{A}}(x_n)$ and $\bar{\mu}_{\tilde{A}}(x_n)$ are the lower and upper memberships of $\tilde{A}$ at x_n.
(iii) Similarly, find K embedded T1 FSs, $\mu_{B_e^k}$ ($k = 1, 2, \ldots, K$), for IT2 FS $\tilde{B}$, i.e.,

$$\mu_{B_e^k}(x_n) = r_k(x_n) \times [\bar{\mu}_{\tilde{B}}(x_n) - \underline{\mu}_{\tilde{B}}(x_n)] + \underline{\mu}_{\tilde{B}}(x_n), \quad n = 1, 2, \ldots, N \tag{11}$$

(iv) Compute an IT2 FS similarity measure $s_M(\tilde{A}, \tilde{B})$ as an average of T1 FS similarity measures s_{hk} that are computed for all of the HK combinations of the embedded T1 FSs for $\tilde{A}$ and $\tilde{B}$, i.e.,

$$s_M(\tilde{A}, \tilde{B}) = \frac{1}{HK} \sum_{h=1}^{H} \sum_{k=1}^{K} s_{hk}, \tag{12}$$

where

$$s_{hk} = s(A_e^h, A_e^k) \tag{13}$$

and s_{hk} can be any T1 FS similarity measure. Jaccard's similarity measure [7]

$$s_J(A, B) = \frac{p(A \cap B)}{p(A \cup B)} = \frac{\int_X \mu_{A\cap B}(x)dx}{\int_X \mu_{A\cup B}(x)dx} \tag{14}$$

is used in this study, where $p(A \cap B)$ and $p(A \cap B)$ are the cardinalities of $A \cap B$ and $A \cup B$, respectively.

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

(i) It does not satisfy reflexivity, i.e., $s_M(\tilde{A}, \tilde{B}) \neq 1$ when $\tilde{A} = \tilde{B}$ because the randomly generated embedded T1 FSs from $\tilde{A}$ and $\tilde{B}$ cannot always be the same.
(ii) It does not satisfy symmetry because of the random numbers.
(iii) $s_M(\tilde{A}, \tilde{B})$ may change from experiment to experiment. When both H and K are large, some kind of stochastic convergence can be expected to occur for $s_M(\tilde{A}, \tilde{B})$ (e.g., convergence in probability); however, the computational cost is heavy because the computation of (12) requires direct enumeration of all HK embedded T1 FSs.

4.2. Gorzalczany's IT2 FS compatibility measure

Gorzalczany [5] defined the *degree of compatibility*, $s_G(\tilde{A}, \tilde{B})$, between two IT2 FSs $\tilde{A}$ and $\tilde{B}$ as

$$s_G(\tilde{A}, \tilde{B}) = \left[\min\left(\frac{\max_{x\in X}\{\min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))\}}{\max_{x\in X} \underline{\mu}_{\tilde{A}}(x)}, \frac{\max_{x\in X}\{\min(\bar{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{B}}(x))\}}{\max_{x\in X} \bar{\mu}_{\tilde{A}}(x)} \right), \right.$$
$$\left. \max\left(\frac{\max_{x\in X}\{\min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))\}}{\max_{x\in X} \underline{\mu}_{\tilde{A}}(x)}, \frac{\max_{x\in X}\{\min(\bar{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{B}}(x))\}}{\max_{x\in X} \bar{\mu}_{\tilde{A}}(x)} \right) \right]. \tag{15}$$

[9] Transitivity and overlapping used in this paper are stronger than their counterparts in [37].

This compatibility measure also does not satisfy reflexivity. It has been shown [37] that as long as $\max_{x\in X}\underline{\mu}_{\widetilde{A}}(x)=\max_{x\in X}\underline{\mu}_{\widetilde{B}}(x)$ and $\max_{x\in X}\bar{\mu}_{\widetilde{A}}(x)=\max_{x\in X}\bar{\mu}_{\widetilde{B}}(x)$, no matter how different the shapes of $\widetilde{A}$ and $\widetilde{B}$ are, this compatibility measure always gives $s_G(\widetilde{A},\widetilde{B})=s_G(\widetilde{B},\widetilde{A})=[1,1]$, which is counter-intuitive.

4.3. Bustince's IT2 FS similarity measure

Bustince's *interval valued normal similarity measure* [3] is defined as

$$s_B(\widetilde{A},\widetilde{B})=[s_L(\widetilde{A},\widetilde{B}),s_U(\widetilde{A},\widetilde{B})] \tag{16}$$

where

$$s_L(\widetilde{A},\widetilde{B})=\Upsilon_L(\widetilde{A},\widetilde{B})\bigstar\Upsilon_L(\widetilde{B},\widetilde{A}) \tag{17}$$

and

$$s_U(\widetilde{A},\widetilde{B})=\Upsilon_U(\widetilde{A},\widetilde{B})\bigstar\Upsilon_U(\widetilde{B},\widetilde{A}), \tag{18}$$

$\bigstar$ 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* of $\widetilde{A}$ in $\widetilde{B}$. $\Upsilon_L(\widetilde{A},\widetilde{B})$ and $\Upsilon_U(\widetilde{A},\widetilde{B})$ used in this study (and taken from [3]) are computed as

$$\Upsilon_L(\widetilde{A},\widetilde{B})=\inf_{x\in X}\left\{1,\min(1-\underline{\mu}_{\widetilde{A}}(x)+\underline{\mu}_{\widetilde{B}}(x),1-\bar{\mu}_{\widetilde{A}}(x)+\bar{\mu}_{\widetilde{B}}(x))\right\} \tag{19}$$

$$\Upsilon_U(\widetilde{A},\widetilde{B})=\inf_{x\in X}\left\{1,\max(1-\underline{\mu}_{\widetilde{A}}(x)+\underline{\mu}_{\widetilde{B}}(x),1-\bar{\mu}_{\widetilde{A}}(x)+\bar{\mu}_{\widetilde{B}}(x))\right\} \tag{20}$$

It has been shown [37] that Bustince's similarity measure does not satisfy overlapping i.e., 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 a nonzero constant, whereas $s_B(\widetilde{A},\widetilde{B})=0$ is expected.

4.4. Zeng and Li's IT2 FS similarity measure

Zeng and Li [45] proposed the following similarity measure for IT2 FSs if the universes of discourse of $\widetilde{A}$ and $\widetilde{B}$ are discrete:

$$s_Z(\widetilde{A},\widetilde{B})=1-\frac{1}{2N}\sum_{i=1}^{N}\left(|\underline{\mu}_{\widetilde{A}}(x_i)-\underline{\mu}_{\widetilde{B}}(x_i)|+|\bar{\mu}_{\widetilde{A}}(x_i)-\bar{\mu}_{\widetilde{B}}(x_i)|\right), \tag{21}$$

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

$$s_Z(\widetilde{A},\widetilde{B})=1-\frac{1}{2(b-a)}\int_a^b(|\underline{\mu}_{\widetilde{A}}(x)-\underline{\mu}_{\widetilde{B}}(x)|+|\bar{\mu}_{\widetilde{A}}(x)-\bar{\mu}_{\widetilde{B}}(x)|)\,dx. \tag{22}$$

A problem [37] with this approach is that when $\widetilde{A}$ and $\widetilde{B}$ are disjoint, the similarity is a nonzero constant, or increases as the distance increases, i.e., it does not satisfy overlapping.

4.5. Vector similarity measure

Recently, Wu and Mendel [37] proposed a vector similarity measure (VSM), which has two components:

$$\mathbf{s}_v(\widetilde{A},\widetilde{B})=(s_1(\widetilde{A},\widetilde{B}),s_2(\widetilde{A},\widetilde{B}))^T \tag{23}$$

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 $\widetilde{B}$.

To compute $s_1(\widetilde{A},\widetilde{B})$, first $c(\widetilde{A})$ and $c(\widetilde{B})$ are computed, and then $\widetilde{B}$ is moved to $\widetilde{B}'$ so that $c(\widetilde{A})=c(\widetilde{B}')$. $s_1(\widetilde{A},\widetilde{B})$ is then computed as the ratio of the *average cardinalities* [see (41)] of $\widetilde{A}\cap\widetilde{B}'$ and $\widetilde{A}\cup\widetilde{B}'$, i.e.

$$\begin{aligned}
s_1(\widetilde{A},\widetilde{B})&\equiv\frac{p(\widetilde{A}\cap\widetilde{B}')}{p(\widetilde{A}\cup\widetilde{B}')}\\
&=\frac{p(\bar{\mu}_{\widetilde{A}}(x)\cap\bar{\mu}_{\widetilde{B}'}(x))+p(\underline{\mu}_{\widetilde{A}}(x)\cap\underline{\mu}_{\widetilde{B}'}(x))}{p(\bar{\mu}_{\widetilde{A}}(x)\cup\bar{\mu}_{\widetilde{B}'}(x))+p(\underline{\mu}_{\widetilde{A}}(x)\cup\underline{\mu}_{\widetilde{B}'}(x))}\\
&=\frac{\int_X\min(\bar{\mu}_{\widetilde{A}}(x),\bar{\mu}_{\widetilde{B}'}(x))dx+\int_X\min(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}'}(x))dx}{\int_X\max(\bar{\mu}_{\widetilde{A}}(x),\bar{\mu}_{\widetilde{B}'}(x))dx+\int_X\max(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}'}(x))\,dx},
\end{aligned} \tag{24}$$

Observe that when all uncertainty disappears, $\widetilde{A}$ and $\widetilde{B}'$ become T1 FSs A and B', and (24) reduces to Jaccard's similarity measure (see (14)).

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

$$s_2(\widetilde{A},\widetilde{B}) = e^{-rd(\widetilde{A},\widetilde{B})}, \tag{25}$$

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

A scalar similarity measure is then computed from the VSM as

$$s_s(\widetilde{A},\widetilde{B}) = s_1(\widetilde{A},\widetilde{B}) \times s_2(\widetilde{A},\widetilde{B}) \tag{26}$$

Though $s_s(\widetilde{A},\widetilde{B})$ decreases as the distance between $\widetilde{A}$ and $\widetilde{B}$ increases, $s_s(\widetilde{A},\widetilde{B})$ does not satisfy overlapping, i.e., when $\widetilde{A}$ and $\widetilde{B}$ are disjoint, $s_s(\widetilde{A},\widetilde{B}) > 0$. This is because:

(i) In $s_1(\widetilde{A},\widetilde{B})$ (see (24)) $\widetilde{B}'$ has the same average centroid as $\widetilde{A}$, and hence $\widetilde{A} \cap \widetilde{B}' \neq \emptyset$, i.e., $s_1(\widetilde{A},\widetilde{B}) > 0$.
(ii) $s_2(\widetilde{A},\widetilde{B})$ is an exponential function, which is always larger than 0.

4.6. The Jaccard similarity measure for IT2 FSs

A new similarity measure, which is an extension of Jaccard's similarity measure for T1 FSs (see (14)), is proposed in this subsection. It is motivated by (24): if $p(\widetilde{A} \cap \widetilde{B})/p(\widetilde{A} \cup \widetilde{B})$ is computed directly instead of $p(\widetilde{A} \cap \widetilde{B}')/p(\widetilde{A} \cup \widetilde{B}')$, then both shape and proximity information are utilized simultaneously without having to align $\widetilde{A}$ and $\widetilde{B}$ and compute their centroids. The new similarity measure is defined as:

$$s_J(\widetilde{A},\widetilde{B}) \equiv \frac{p(\widetilde{A} \cap \widetilde{B})}{p(\widetilde{A} \cup \widetilde{B})} = \frac{\int_X \min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x))\, dx + \int_X \min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\, dx}{\int_X \max(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) dx + \int_X \max(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\, dx}. \tag{27}$$

Theorem 2. *The Jaccard similarity measure satisfies reflexivity, symmetry, transitivity and overlapping.*

Proof 2. The four properties are proved in order next.

[P1.] *Reflexivity:* Consider first the necessity, i.e., $s_J(\widetilde{A},\widetilde{B}) = 1 \Rightarrow \widetilde{A} = \widetilde{B}$. When the areas of the FOUs are not zero, $\min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x)) < \max(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x))$; hence, the only way that $s_J(\widetilde{A},\widetilde{B}) = 1$ (see (27)) is when $\min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) = \max(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x))$ and $\min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x)) = \max(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))$, in which case $\bar{\mu}_{\widetilde{A}}(x) = \bar{\mu}_{\widetilde{B}}(x)$ and $\underline{\mu}_{\widetilde{A}}(x) = \underline{\mu}_{\widetilde{B}}(x)$, i.e., $\widetilde{A} = \widetilde{B}$. Consider next the sufficiency, i.e., $\widetilde{A} = \widetilde{B} \Rightarrow s_J(\widetilde{A},\widetilde{B}) = 1$. When $\widetilde{A} = \widetilde{B}$, i.e., $\bar{\mu}_{\widetilde{A}}(x) = \bar{\mu}_{\widetilde{B}}(x)$ and $\underline{\mu}_{\widetilde{A}}(x) = \underline{\mu}_{\widetilde{B}}(x)$, it follows that $\min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) = \max(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x))$ and $\min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x)) = \max(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))$. Consequently, it follows from (27) that $s_J(\widetilde{A},\widetilde{B}) = 1$.
[P2.] *Symmetry:* Observe from (27) that $s_J(\widetilde{A},\widetilde{B})$ does not depend on the order of $\widetilde{A}$ and $\widetilde{B}$; so, $s_J(\widetilde{A},\widetilde{B}) = s_J(\widetilde{B},\widetilde{A})$.
[P3.] *Transitivity:* If $\widetilde{A} \leqslant \widetilde{B} \leqslant \widetilde{C}$ (see Definition 3), then

$$\begin{aligned} s_J(\widetilde{A},\widetilde{B}) &= \frac{\int_X \min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) dx + \int_X \min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x)) dx}{\int_X \max(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) dx + \int_X \max(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x)) dx} \\ &= \frac{\int_X \bar{\mu}_{\widetilde{A}}(x) dx + \int_X \underline{\mu}_{\widetilde{A}}(x) dx}{\int_X \bar{\mu}_{\widetilde{B}}(x) dx + \int_X \underline{\mu}_{\widetilde{B}}(x) dx} \end{aligned} \tag{28}$$

$$\begin{aligned} s_J(\widetilde{A},\widetilde{C}) &= \frac{\int_X \min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{C}}(x)) dx + \int_X \min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{C}}(x)) dx}{\int_X \max(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{C}}(x)) dx + \int_X \max(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{C}}(x)) dx} \\ &= \frac{\int_X \bar{\mu}_{\widetilde{A}}(x) dx + \int_X \underline{\mu}_{\widetilde{A}}(x) dx}{\int_X \bar{\mu}_{\widetilde{C}}(x) dx + \int_X \underline{\mu}_{\widetilde{C}}(x) dx} \end{aligned} \tag{29}$$

Because $\widetilde{B} \leqslant \widetilde{C}$, it follows that $\int_X \bar{\mu}_{\widetilde{B}}(x) dx + \int_X \underline{\mu}_{\widetilde{B}}(x) dx \leqslant \int_X \bar{\mu}_{\widetilde{C}}(x) dx + \int_X \underline{\mu}_{\widetilde{C}}(x) dx$, and hence $s_J(\widetilde{A},\widetilde{B}) \geqslant s_J(\widetilde{A},\widetilde{C})$.
[P4.] *Overlapping:* If $\widetilde{A} \cap \widetilde{B} \neq \emptyset$ (see Definition 1), $\exists x$ such that $\min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) > 0$, then, in the numerator of (27),

$$\int_X \min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) dx + \int_X \min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\, dx > 0 \tag{30}$$

In the denominator of (27),

$$\begin{aligned} &\int_X \max(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x))\, dx + \int_X \max(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\, dx \\ &\geqslant \int_X \min(\bar{\mu}_{\widetilde{A}}(x), \bar{\mu}_{\widetilde{B}}(x)) dx + \int_X \min(\underline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{B}}(x))\, dx > 0 \end{aligned} \tag{31}$$

Consequently, $s_J(\widetilde{A},\widetilde{B}) > 0$. On the other hand, when $\widetilde{A}\cap\widetilde{B}=\emptyset$, i.e., $\min(\bar{\mu}_{\widetilde{A}}(x),\bar{\mu}_{\widetilde{B}}(x)) = \min(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}}(x)) = 0$ for $\forall x$, then, in the numerator of (27),

$$\int_X \min(\bar{\mu}_{\widetilde{A}}(x),\bar{\mu}_{\widetilde{B}}(x))\,dx + \int_X \min(\underline{\mu}_{\widetilde{A}}(x),\underline{\mu}_{\widetilde{B}}(x))\,dx = 0 \tag{32}$$

Consequently, $s_J(\widetilde{A},\widetilde{B}) = 0$. □

4.7. Comparative studies

We have shown that the Jaccard similarity measure satisfies all four desirable properties of a similarity measure. Next, the performances of the six similarity measure are compared using the 32 word FOUs depicted in Fig. 2. The similarities are summarized in Tables 2–7 respectively. Each table contains a matrix of 1024 entries, so we shall guide the reader next to their critical highlights. Observe that:

(i) Table 2: Examining the diagonal elements of this table, we see that Mitchell's method gives $s_M(\widetilde{A},\widetilde{A}) < 1$. Also, because $s_M(\widetilde{A},\widetilde{B}) \neq s_M(\widetilde{B},\widetilde{A})$, the matrix is not symmetric.
(ii) Table 3: Examining the block of ones at the bottom-right corner of this table, we see that Gorzalczany's method indicates "very large (27)," "humongous amount (28)," "huge amount (29)," "very high amount (30)," "extreme amount (31)" and "maximum amount (32)" are equivalent, which is counter-intuitive because their FOUs are not completely the same (see Fig. 2).
(iii) Table 4: Examining element (6,7) of this table, we see that Bustince's method shows the similarity between "very little" and "a bit" is zero, and examining element (26,27), we see that the similarity between "large" and "very large" is also zero, both of which are counter-intuitive.
(iv) Table 5: Examining this table, we see that all similarities are larger than 0.50, i.e., Zeng and Li's method gives large similarity whether or not $\widetilde{A}$ and $\widetilde{B}$ overlap. Examining the first line of this table, we see that the similarity generally decreases and then increases as two words get further away, whereas a monotonically decreasing trend is expected.
(v) Table 6: Examining this table, we see that the VSM gives very reasonable results. Generally the similarity decreases monotonically as two words gets further away[10]. Note also that there are zeros in the table because only two digits are used. Theoretically $s_J(\widetilde{A},\widetilde{B})$ is always larger than zero (see the arguments under (26)).
(vi) Table 7: Comparing this table with Table 6, we see that Jaccard's similarity measure gives similar results to the VSM, but they are more reliable (e.g., the zeros are true zeros instead of the results of roundoff). Also, simulations show that Jaccard's method is about 3.5 times faster than the VSM.
(vii) Except for Mitchell's method, all other similarity measures indicate that "sizeable (19)" and "quite a bit (20)" are equivalent, and "high amount (23)" and "substantial amount (24)" are equivalent (i.e., their similarities equal 1), which seems reasonable because Table 1 shows that the FOUs of "sizeable" and "quite a bit" are exactly the same, and the FOUs of "high amount" and "substantial amount" are also exactly the same.

These results suggest that Jaccard's similarity measure should be used for CWW.

It is also interesting to know which words are similar to a particular word with similarity values larger than a pre-specified threshold. When the Jaccard similarity measure is used, the groups of similar words for different thresholds are shown in Table 8, e.g., Row 1 shows that the words "teeny-weeny (2)," "a smidgen (3)" and "tiny (4)" are similar to the word "none to very little (1)" to degree $\geqslant 0.7$, and that these three words as wells as the words "very small (5)" and "very little (6)" are similar to "none to very little (1)" to degree $\geqslant 0.6$. Observe that except for the word "maximum amount (32)," every word in the 32 word vocabulary has at least one word similar to it with similarity larger than or equal to 0.6. Observe, also, that there are five words [considerable amount (21), substantial amount (22), a lot (23), high amount (24), and very sizeable (25)] with the most number (7 in this example) of neighbors with similarity larger than or equal to 0.5, and all of them have interior FOUs (see Fig. 2). The fact that so many of the 32 words are similar to many other words suggest that it is possible to create many sub-vocabularies that cover the interval $[0,10]$. Some examples of five word vocabularies are given in [11].

5. Uncertainty measures

Wu and Mendel [34] proposed five uncertainty measures for IT2 FSs: *centroid, cardinality, fuzziness, variance* and *skewness*; however, an open question is which one to use. In this section, this question is tackled by distinguishing between *intra-personal uncertainty* and *inter-personal uncertainty* [29], and studying which uncertainty measure best captures both of them.

[10] There are cases where the similarity does not decrease monotonically, e.g., elements 4 and 5 in the first row. This is because the distances among the words are determined by a ranking method which considers only the centroids but not the shapes of the IT2 FSs. Additional discussions are given in the last paragraph of this subsection.

Table 2
Similarity matrix when Mitchell's similarity measure is used.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. None to very little	.71	.57	.62	.61	.60	.60	.11	.09	.13	.11	.06	.04	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. Teeny-weeny	.57	.54	.50	.48	.46	.46	.12	.10	.13	.11	.07	.04	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. A smidgen	.62	.50	.65	.65	.66	.65	.18	.16	.20	.17	.11	.07	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.60	.49	.65	.65	.66	.66	.20	.16	.20	.17	.12	.07	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very small	.60	.49	.64	.66	.75	.72	.18	.16	.19	.17	.10	.07	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very little	.59	.47	.65	.65	.72	.71	.20	.17	.21	.18	.11	.08	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.11	.11	.19	.19	.18	.20	.70	.50	.44	.42	.36	.17	.09	.02	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8. Little	.09	.10	.16	.17	.15	.17	.51	.58	.52	.53	.49	.29	.17	.08	.10	.03	.02	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9. Low amount	.13	.13	.19	.19	.20	.22	.44	.53	.51	.52	.47	.27	.17	.08	.09	.03	.02	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10. Small	.11	.11	.16	.17	.16	.18	.42	.52	.52	.54	.50	.31	.19	.09	.10	.03	.03	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11. Somewhat small	.07	.07	.11	.11	.10	.11	.36	.48	.45	.50	.55	.30	.21	.09	.11	.03	.02	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12. Some	.04	.04	.07	.07	.07	.08	.19	.27	.26	.31	.30	.53	.49	.38	.39	.25	.29	.21	.13	.13	.13	.06	.06	.07	.06	.04	0	0	0	0	0	0
13. Some to moderate	.01	.01	.03	.03	.02	.03	.09	.17	.17	.19	.21	.49	.54	.46	.47	.32	.38	.27	.16	.15	.16	.08	.07	.08	.08	.05	0	0	0	0	0	0
14. Moderate amount	0	0	0	0	0	0	.02	.08	.08	.10	.09	.39	.47	.55	.52	.40	.44	.30	.18	.17	.16	.08	.08	.08	.08	.05	0	0	0	0	0	0
15. Fair amount	0	0	0	0	0	0	.03	.10	.09	.10	.11	.39	.48	.51	.55	.38	.44	.34	.22	.21	.20	.10	.09	.09	.10	.06	0	0	0	0	0	0
16. Medium	0	0	0	0	0	0	0	.03	.03	.03	.03	.25	.33	.40	.39	.51	.48	.27	.15	.16	.15	.06	.05	.06	.06	.02	0	0	0	0	0	0
17. Modest amount	0	0	0	0	0	0	0	.02	.02	.03	.02	.28	.37	.44	.46	.48	.55	.35	.20	.20	.20	.09	.09	.08	.09	.05	0	0	0	0	0	0
18. Good amount	0	0	0	0	0	0	0	.02	.02	.02	.02	.21	.27	.30	.35	.27	.35	.55	.46	.48	.46	.32	.32	.31	.31	.27	.08	.08	.08	.08	.08	.02
19. Sizeable	0	0	0	0	0	0	0	0	0	0	0	.13	.16	.17	.20	.15	.22	.48	.57	.58	.54	.42	.43	.41	.40	.37	.10	.09	.09	.09	.09	.02
20. Quite a bit	0	0	0	0	0	0	0	0	0	0	0	.13	.16	.17	.20	.15	.21	.48	.58	.58	.54	.41	.43	.44	.41	.37	.10	.10	.09	.09	.10	.02
21. Considerable amount	0	0	0	0	0	0	0	0	0	0	0	.12	.16	.17	.20	.15	.20	.45	.54	.54	.54	.44	.44	.45	.42	.39	.12	.12	.12	.12	.12	.03
22. Substantial amount	0	0	0	0	0	0	0	0	0	0	0	.06	.08	.08	.10	.06	.09	.32	.43	.43	.45	.57	.56	.57	.54	.52	.20	.19	.18	.19	.17	.06
23. A lot	0	0	0	0	0	0	0	0	0	0	0	.06	.08	.08	.10	.06	.09	.31	.43	.44	.45	.56	.57	.58	.53	.52	.20	.19	.20	.18	.19	.06
24. High amount	0	0	0	0	0	0	0	0	0	0	0	.06	.07	.08	.10	.05	.09	.31	.41	.41	.45	.57	.58	.59	.57	.55	.21	.20	.19	.19	.18	.06
25. Very sizeable	0	0	0	0	0	0	0	0	0	0	0	.06	.08	.08	.10	.06	.09	.31	.40	.40	.43	.54	.54	.56	.53	.50	.23	.21	.22	.20	.20	.07
26. Large	0	0	0	0	0	0	0	0	0	0	0	.04	.05	.05	.07	.03	.05	.26	.37	.37	.41	.53	.53	.55	.50	.61	.19	.19	.18	.18	.18	.04
27. Very large	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.08	.10	.10	.12	.20	.20	.21	.23	.18	.76	.56	.74	.66	.64	.39
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.08	.10	.10	.11	.19	.19	.20	.20	.19	.55	.57	.59	.57	.58	.43
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.08	.10	.09	.11	.19	.20	.20	.21	.18	.74	.58	.73	.68	.66	.43
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.07	.09	.10	.11	.18	.19	.19	.20	.18	.66	.58	.68	.66	.66	.46
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.08	.09	.09	.11	.18	.18	.19	.20	.18	.65	.58	.66	.66	.66	.46
32. Maximum amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.02	.03	.06	.06	.06	.07	.04	.40	.44	.43	.46	.47	.69

Table 3
Similarity matrix when Gorzalczany's similarity measure is used.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. None to very little	1	1	1	1	1	1	.25	.27	.31	.29	.21	.20	.09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. Teeny-weeny	1	1	.99	.99	.99	.99	.25	.27	.31	.29	.21	.20	.09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. A smidgen	1	.99	1	1	1	1	.31	.32	.36	.34	.27	.25	.15	0	.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	1	.99	1	1	1	1	.32	.33	.37	.34	.28	.25	.16	0	.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very small	1	.99	1	1	1	1	.42	.37	.40	.37	.29	.26	.15	0	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very little	1	.99	1	1	1	1	.40	.36	.40	.37	.30	.27	.17	0	.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.25	.25	.31	.32	.45	.41	1	.85	.70	.64	.50	.43	.30	.14	.18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8. Little	.27	.27	.32	.33	.38	.36	.99	.99	.78	.70	.50	.50	.39	.28	.29	.15	.15	.14	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0
9. Low amount	.31	.31	.36	.37	.40	.40	.99	.99	.99	.81	.50	.50	.38	.29	.29	.15	.15	.15	.02	.02	.02	0	0	0	0	0	0	0	0	0	0	0
10. Small	.29	.29	.34	.34	.37	.37	.76	.76	.73	.99	.78	.50	.43	.33	.33	.18	.18	.16	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0
11. Somewhat small	.21	.21	.27	.28	.29	.30	.50	.50	.50	.77	.99	.61	.46	.35	.34	.18	.18	.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12. Some	.20	.20	.25	.25	.26	.27	.43	.50	.50	.50	.63	.99	.55	.50	.50	.50	.50	.45	.35	.35	.34	.24	.25	.24	.24	.20	.04	.04	.04	.04	.04	0
13. Some to moderate	.09	.09	.15	.16	.15	.17	.30	.39	.38	.43	.46	.55	.99	.95	.75	.77	.75	.50	.39	.39	.39	.27	.27	.27	.27	.22	.05	.05	.05	.05	.05	0
14. Moderate amount	0	0	0	0	0	0	.14	.28	.29	.33	.35	.50	.97	1	.73	.75	.72	.50	.38	.38	.38	.26	.26	.26	.26	.21	.03	.02	.02	.02	.02	0
15. Fair amount	0	0	.05	.05	.03	.05	.18	.29	.29	.33	.34	.50	.74	.71	.99	.75	.99	.57	.44	.44	.44	.31	.31	.31	.31	.25	.06	.05	.05	.05	.05	0
16. Medium	0	0	0	0	0	0	0	.15	.15	.18	.18	.50	.90	.85	.88	.98	.88	.50	.36	.36	.35	.22	.21	.22	.22	.15	0	0	0	0	0	0
17. Modest amount	0	0	0	0	0	0	0	.15	.15	.18	.18	.50	.74	.70	.98	.75	.99	.50	.43	.43	.43	.28	.28	.28	.28	.22	0	0	0	0	0	0
18. Good amount	0	0	0	0	0	0	0	.14	.15	.16	.15	.45	.50	.50	.57	.50	.50	1	.67	.67	.50	.50	.50	.50	.50	.46	.27	.25	.26	.25	.25	.14
19. Sizeable	0	0	0	0	0	0	0	.03	.02	.03	0	.35	.39	.38	.44	.36	.43	.64	.99	.99	.87	.66	.69	.66	.50	.58	.29	.26	.28	.27	.26	.14
20. Quite a bit	0	0	0	0	0	0	0	.03	.02	.03	0	.35	.39	.38	.44	.36	.43	.64	.99	.99	.87	.66	.69	.66	.50	.58	.29	.26	.28	.27	.26	.14
21. Considerable amount	0	0	0	0	0	0	0	.03	.02	.03	0	.34	.39	.38	.44	.35	.43	.50	1	1	1	.71	.75	.71	.50	.60	.33	.29	.31	.30	.29	.18
22. Substantial amount	0	0	0	0	0	0	0	0	0	0	0	.24	.27	.26	.31	.22	.28	.50	.67	.67	.67	.99	.99	.99	.78	.95	.42	.35	.37	.36	.35	.25
23. A lot	0	0	0	0	0	0	0	0	0	0	0	.25	.27	.26	.31	.21	.28	.50	.70	.70	.70	.97	.99	.97	.77	.93	.42	.35	.37	.36	.35	.25
24. High amount	0	0	0	0	0	0	0	0	0	0	0	.24	.27	.26	.31	.22	.28	.50	.64	.64	.64	.92	.92	1	.85	.98	.54	.35	.45	.36	.35	.25
25. Very sizeable	0	0	0	0	0	0	0	0	0	0	0	.24	.27	.26	.31	.22	.28	.50	.50	.50	.50	.83	.84	.99	.99	.99	.64	.37	.52	.38	.37	.27
26. Large	0	0	0	0	0	0	0	0	0	0	0	.20	.22	.21	.25	.15	.22	.46	.57	.57	.57	.85	.86	.95	.83	.99	.51	.32	.42	.33	.32	.19
27. Very large	0	0	0	0	0	0	0	0	0	0	0	.04	.05	.03	.06	0	0	.27	.29	.29	.33	.40	.41	.47	.51	.44	1	1	1	1	1	1
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	.04	.05	.02	.05	0	0	.25	.26	.26	.29	.35	.35	.35	.37	.32	1	1	1	1	1	1
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	.04	.05	.02	.05	0	0	.26	.28	.28	.31	.37	.37	.41	.45	.38	1	1	1	1	1	1
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	.04	.05	.02	.05	0	0	.25	.27	.27	.30	.36	.36	.36	.38	.33	1	1	1	1	1	1
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	.04	.05	.02	.05	0	0	.25	.26	.26	.29	.35	.35	.35	.37	.32	1	1	1	1	1	1
32. Maximum amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.14	.14	.14	.18	.25	.25	.25	.27	.19	1	1	1	1	1	1

Table 4
Similarity matrix when Bustince's similarity measure is used.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. None to very little	1	.57	.86	.86	.63	.68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. Teeny-weeny	.57	1	.52	.50	.38	.38	.03	.05	.05	.05	.05	.12	.12	.12	.12	.10	.11	.16	.15	.15	.16	.16	.16	.16	.16	.16	0	.06	0	0	0	.03
3. A smidgen	.86	.52	1	.98	.67	.72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.86	.50	.98	1	.69	.74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very small	.63	.38	.67	.69	1	.92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very little	.68	.38	.72	.74	.92	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	0	.03	0	0	0	0	1	.72	.64	.48	.35	.26	.13	.13	.13	.14	.13	.13	.14	.14	.13	.14	.14	.14	.14	.14	0	.06	0	0	0	.02
8. Little	0	.05	0	0	0	0	.72	1	.79	.71	.58	.30	.24	.24	.24	.24	.24	.24	.24	.24	.24	.24	.25	.24	.24	.22	0	.06	0	0	0	.03
9. Low amount	0	.05	0	0	0	0	.64	.79	1	.76	.63	.33	.30	.31	.30	.36	.30	.30	.26	.26	.32	.28	.27	.24	.32	.22	0	.06	0	0	0	.03
10. Small	0	.05	0	0	0	0	.48	.71	.76	1	.77	.33	.30	.30	.30	.30	.30	.30	.26	.26	.30	.28	.27	.24	.30	.22	0	.06	0	0	0	.03
11. Somewhat small	0	.05	0	0	0	0	.35	.58	.63	.77	1	.34	.30	.29	.30	.29	.29	.29	.26	.26	.29	.28	.27	.24	.29	.22	0	.06	0	0	0	.02
12. Some	0	.12	0	0	0	0	.26	.30	.33	.33	.34	1	.72	.34	.42	.33	.33	.30	.26	.26	.32	.28	.27	.24	.32	.22	0	.06	0	0	0	.03
13. Some to moderate	0	.12	0	0	0	0	.13	.24	.30	.30	.30	.72	1	.70	.72	.38	.38	.32	.26	.26	.31	.28	.27	.24	.31	.22	0	.05	0	0	0	.02
14. Moderate amount	0	.12	0	0	0	0	.13	.24	.31	.30	.29	.34	.70	1	.73	.55	.55	.33	.26	.26	.32	.28	.27	.24	.31	.22	0	.04	0	0	0	.01
15. Fair amount	0	.12	0	0	0	0	.13	.24	.30	.30	.30	.42	.72	.73	1	.58	.66	.41	.26	.26	.30	.28	.27	.24	.30	.22	0	.04	0	0	0	.01
16. Medium	0	.10	0	0	0	0	.14	.24	.36	.30	.29	.33	.38	.55	.58	1	.75	.30	.26	.26	.32	.28	.27	.24	.31	.22	0	.03	0	0	0	0
17. Modest amount	0	.11	0	0	0	0	.13	.24	.30	.30	.29	.33	.38	.55	.66	.75	1	.31	.26	.26	.29	.28	.27	.24	.29	.22	0	.03	0	0	0	0
18. Good amount	0	.16	0	0	0	0	.13	.24	.30	.30	.29	.30	.32	.33	.41	.30	.31	1	.67	.67	.67	.32	.33	.31	.32	.26	0	.03	0	0	0	0
19. Sizeable	0	.15	0	0	0	0	.14	.24	.26	.26	.26	.26	.26	.26	.26	.26	.26	.67	1	1	.84	.59	.59	.54	.46	.42	0	.02	0	0	0	0
20. Quite a bit	0	.15	0	0	0	0	.14	.24	.26	.26	.26	.26	.26	.26	.26	.26	.26	.67	1	1	.84	.59	.59	.54	.46	.42	0	.02	0	0	0	0
21. Considerable amount	0	.16	0	0	0	0	.13	.24	.32	.30	.29	.32	.31	.32	.30	.32	.29	.67	.84	.84	1	.66	.66	.59	.57	.50	0	.02	0	0	0	0
22. Substantial amount	0	.16	0	0	0	0	.14	.24	.28	.28	.28	.28	.28	.28	.28	.28	.28	.32	.59	.59	.66	1	.95	.88	.70	.86	0	.01	0	0	0	0
23. A lot	0	.16	0	0	0	0	.14	.25	.27	.27	.27	.27	.27	.27	.27	.27	.27	.33	.59	.59	.66	.95	1	.88	.69	.83	0	.01	0	0	0	0
24. High amount	0	.16	0	0	0	0	.14	.24	.24	.24	.24	.24	.24	.24	.24	.24	.24	.31	.54	.54	.59	.88	.88	1	.74	.85	0	.01	0	0	0	0
25. Very sizeable	0	.16	0	0	0	0	.14	.24	.32	.30	.29	.32	.31	.31	.30	.31	.29	.32	.46	.46	.57	.70	.69	.74	1	.75	0	.01	0	0	0	0
26. Large	0	.16	0	0	0	0	.14	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22	.26	.42	.42	.50	.86	.83	.85	.75	1	0	0	0	0	0	0
27. Very large	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	.49	.86	.67	.64	.40
28. Humongous amount	0	.06	0	0	0	0	.06	.06	.06	.06	.06	.06	.05	.04	.04	.03	.03	.03	.02	.02	.02	.01	.01	.01	.01	0	.49	1	.55	.66	.68	.73
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.86	.55	1	.77	.74	.49
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.67	.66	.77	1	.96	.60
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.64	.68	.74	.96	1	.63
32. Maximum amount	0	.03	0	0	0	0	.02	.03	.03	.03	.02	.03	.02	.01	.01	0	0	0	0	0	0	0	0	0	0	0	.40	.73	.49	.60	.63	1

Table 5
Similarity matrix when Zeng and Li's similarity measure is used.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. None to very little	1	.93	.92	.91	.84	.86	.58	.63	.66	.62	.60	.62	.63	.69	.66	.75	.72	.70	.74	.74	.74	.78	.77	.77	.77	.80	.82	.86	.83	.84	.85	.89
2. Teeny-weeny	.93	1	.88	.87	.79	.80	.62	.66	.69	.65	.63	.64	.65	.70	.68	.77	.74	.72	.75	.75	.75	.79	.79	.78	.79	.81	.83	.87	.85	.86	.86	.90
3. A smidgen	.92	.88	1	.99	.92	.93	.61	.65	.69	.65	.62	.63	.63	.66	.64	.72	.70	.68	.72	.72	.72	.76	.75	.75	.75	.78	.80	.84	.81	.83	.83	.87
4. Tiny	.91	.87	.99	1	.92	.94	.61	.65	.69	.65	.62	.63	.62	.66	.64	.72	.70	.68	.71	.71	.72	.75	.75	.75	.75	.77	.80	.84	.81	.82	.83	.87
5. Very small	.84	.79	.92	.92	1	.97	.56	.61	.65	.61	.57	.60	.60	.64	.61	.70	.67	.66	.70	.70	.70	.74	.73	.73	.73	.76	.78	.82	.79	.81	.81	.85
6. Very little	.86	.80	.93	.94	.97	1	.58	.63	.67	.63	.59	.61	.60	.64	.62	.70	.67	.66	.70	.70	.70	.74	.73	.73	.73	.76	.78	.82	.79	.81	.81	.85
7. A bit	.58	.62	.61	.61	.56	.58	1	.86	.82	.78	.73	.68	.63	.62	.61	.66	.64	.64	.67	.67	.68	.72	.71	.71	.71	.74	.77	.81	.78	.79	.79	.84
8. Little	.63	.66	.65	.65	.61	.63	.86	1	.95	.91	.86	.74	.68	.64	.64	.64	.62	.62	.64	.64	.65	.68	.68	.68	.68	.70	.73	.77	.74	.76	.76	.80
9. Low amount	.66	.69	.69	.69	.65	.67	.82	.95	1	.94	.87	.75	.69	.65	.65	.66	.63	.63	.65	.65	.65	.69	.69	.68	.69	.71	.74	.78	.75	.76	.77	.81
10. Small	.62	.65	.65	.65	.61	.63	.78	.91	.94	1	.90	.76	.69	.65	.64	.64	.62	.62	.64	.64	.64	.68	.68	.67	.67	.70	.73	.76	.74	.75	.75	.79
11. Somewhat small	.60	.63	.62	.62	.57	.59	.73	.86	.87	.90	1	.79	.70	.65	.65	.64	.62	.62	.64	.64	.65	.69	.69	.68	.68	.71	.74	.78	.75	.76	.77	.81
12. Some	.62	.64	.63	.63	.60	.61	.68	.74	.75	.76	.79	1	.89	.84	.82	.77	.76	.70	.65	.65	.65	.63	.63	.62	.62	.62	.62	.66	.63	.64	.64	.69
13. Some to moderate	.63	.65	.63	.62	.60	.60	.63	.68	.69	.69	.70	.89	1	.91	.89	.81	.82	.73	.67	.67	.67	.64	.63	.63	.63	.63	.62	.66	.63	.64	.65	.69
14. Moderate amount	.69	.70	.66	.66	.64	.64	.62	.64	.65	.65	.65	.84	.91	1	.92	.86	.86	.74	.67	.67	.67	.63	.63	.62	.63	.62	.62	.67	.63	.65	.66	.71
15. Fair amount	.66	.68	.64	.64	.61	.62	.61	.64	.65	.64	.65	.82	.89	.92	1	.83	.89	.77	.69	.69	.69	.64	.64	.63	.64	.63	.61	.66	.62	.64	.64	.69
16. Medium	.75	.77	.72	.72	.70	.70	.66	.64	.66	.64	.64	.77	.81	.86	.83	1	.91	.76	.67	.67	.67	.64	.64	.63	.64	.64	.66	.72	.68	.70	.71	.77
17. Modest amount	.72	.74	.70	.70	.67	.67	.64	.62	.63	.62	.62	.76	.82	.86	.89	.91	1	.79	.68	.68	.69	.63	.63	.62	.62	.61	.61	.67	.63	.65	.66	.72
18. Good amount	.70	.72	.68	.68	.66	.66	.64	.62	.63	.62	.62	.70	.73	.74	.77	.76	.79	1	.86	.86	.85	.75	.75	.74	.74	.72	.59	.64	.60	.62	.62	.62
19. Sizeable	.74	.75	.72	.71	.70	.70	.67	.64	.65	.64	.64	.65	.67	.67	.69	.67	.68	.86	1	1	.96	.81	.81	.80	.80	.76	.59	.64	.60	.62	.62	.62
20. Quite a bit	.74	.75	.72	.71	.70	.70	.67	.64	.65	.64	.64	.65	.67	.67	.69	.67	.68	.86	1	1	.96	.81	.81	.80	.80	.76	.59	.64	.60	.62	.62	.62
21. Considerable amount	.74	.75	.72	.72	.70	.70	.68	.65	.65	.64	.65	.65	.67	.67	.69	.67	.69	.85	.96	.96	1	.84	.84	.83	.83	.80	.60	.66	.62	.64	.64	.63
22. Substantial amount	.78	.79	.76	.75	.74	.74	.72	.68	.69	.68	.69	.63	.64	.63	.64	.64	.63	.75	.81	.81	.84	1	.99	.98	.96	.90	.63	.69	.64	.67	.67	.63
23. A lot	.77	.79	.75	.75	.73	.73	.71	.68	.69	.68	.69	.63	.63	.63	.64	.64	.63	.75	.81	.81	.84	.99	1	.98	.95	.90	.63	.69	.64	.66	.66	.63
24. High amount	.77	.78	.75	.75	.73	.73	.71	.68	.68	.67	.68	.62	.63	.62	.63	.63	.62	.74	.80	.80	.83	.98	.98	1	.96	.92	.62	.67	.63	.65	.65	.62
25. Very sizeable	.77	.79	.75	.75	.73	.73	.71	.68	.69	.67	.68	.62	.63	.63	.64	.64	.62	.74	.80	.80	.83	.96	.95	.96	1	.90	.65	.69	.65	.67	.67	.63
26. Large	.80	.81	.78	.77	.76	.76	.74	.70	.71	.70	.71	.62	.63	.62	.63	.64	.61	.72	.76	.76	.80	.90	.90	.92	.90	1	.61	.67	.62	.64	.64	.62
27. Very large	.82	.83	.80	.80	.78	.78	.77	.73	.74	.73	.74	.62	.62	.62	.61	.66	.61	.59	.59	.59	.60	.63	.63	.62	.65	.61	1	.86	.96	.91	.90	.74
28. Humongous amount	.86	.87	.84	.84	.82	.82	.81	.77	.78	.76	.78	.66	.66	.67	.66	.72	.67	.64	.64	.64	.66	.69	.69	.67	.69	.67	.86	1	.90	.95	.96	.85
29. Huge amount	.83	.85	.81	.81	.79	.79	.78	.74	.75	.74	.75	.63	.63	.63	.62	.68	.63	.60	.60	.60	.62	.64	.64	.63	.65	.62	.96	.90	1	.95	.94	.78
30. Very high amount	.84	.86	.83	.82	.81	.81	.79	.76	.76	.75	.76	.64	.64	.65	.64	.70	.65	.62	.62	.62	.64	.67	.66	.65	.67	.64	.91	.95	.95	1	.99	.83
31. Extreme amount	.85	.86	.83	.83	.81	.81	.79	.76	.77	.75	.77	.64	.65	.66	.64	.71	.66	.62	.62	.62	.64	.67	.66	.65	.67	.64	.90	.96	.94	.99	1	.84
32. Maximum amount	.89	.90	.87	.87	.85	.85	.84	.80	.81	.79	.81	.69	.69	.71	.69	.77	.72	.62	.62	.62	.63	.63	.63	.62	.63	.62	.74	.85	.78	.83	.84	1

Table 6
Similarity matrix when the VSM [37] is used.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. None to very little	1	.54	.51	.49	.48	.47	.09	.08	.08	.07	.04	.04	.02	.01	.01	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. Teeny-weeny	.54	1	.57	.54	.44	.44	.08	.08	.08	.07	.04	.03	.02	.01	.01	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. A smidgen	.51	.57	1	.96	.76	.78	.15	.13	.12	.10	.07	.05	.03	.01	.01	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.49	.54	.96	1	.79	.81	.15	.14	.12	.10	.07	.05	.03	.01	.01	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very small	.48	.44	.76	.79	1	.91	.17	.14	.12	.11	.07	.05	.03	.01	.02	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very little	.47	.44	.78	.81	.91	1	.18	.15	.13	.12	.08	.06	.03	.02	.02	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.09	.08	.15	.15	.17	.18	1	.43	.35	.32	.25	.11	.07	.04	.04	.01	.01	.02	.01	.01	.01	0	0	0	0	0	0	0	0	0	0	0
8. Little	.08	.08	.13	.14	.14	.15	.43	1	.77	.66	.50	.21	.13	.08	.08	.04	.04	.04	.01	.01	.01	.01	.01	.01	.01	0	0	0	0	0	0	0
9. Low amount	.08	.08	.12	.12	.12	.13	.35	.77	1	.80	.55	.23	.15	.10	.09	.05	.05	.04	.02	.02	.02	.01	.01	.01	.01	0	0	0	0	0	0	0
10. Small	.07	.07	.10	.10	.11	.12	.32	.66	.80	1	.64	.25	.18	.11	.11	.05	.05	.05	.02	.02	.02	.01	.01	.01	.01	0	0	0	0	0	0	0
11. Somewhat small	.04	.04	.07	.07	.07	.08	.25	.50	.55	.64	1	.24	.18	.11	.11	.05	.05	.05	.02	.02	.02	.01	.01	.01	.01	0	0	0	0	0	0	0
12. Some	.04	.03	.05	.05	.05	.06	.11	.21	.23	.25	.24	1	.58	.37	.36	.20	.23	.20	.11	.11	.11	.06	.06	.06	.06	.04	.02	.01	.02	.01	.01	.01
13. Some to moderate	.02	.02	.03	.03	.03	.03	.07	.13	.15	.18	.18	.58	1	.57	.60	.31	.34	.29	.16	.16	.16	.09	.09	.08	.08	.06	.02	.02	.02	.02	.02	.01
14. Moderate amount	.01	.01	.01	.01	.01	.02	.04	.08	.10	.11	.11	.37	.57	1	.72	.50	.54	.29	.16	.16	.15	.08	.08	.07	.07	.05	.01	.01	.01	.01	.01	0
15. Fair amount	.01	.01	.01	.01	.02	.02	.04	.08	.09	.11	.11	.36	.60	.72	1	.50	.53	.36	.21	.21	.20	.11	.11	.10	.10	.07	.02	.02	.02	.02	.02	.01
16. Medium	0	0	0	0	0	0	.01	.04	.05	.05	.05	.20	.31	.50	.50	1	.61	.20	.12	.12	.11	.06	.06	.05	.05	.03	.01	.01	.01	.01	.01	0
17. Modest amount	0	0	0	0	0	0	.01	.04	.05	.05	.05	.23	.34	.54	.53	.61	1	.30	.18	.18	.16	.09	.09	.08	.08	.05	.01	.01	.01	.01	.01	0
18. Good amount	.01	.01	.01	.01	.01	.01	.02	.04	.04	.05	.05	.20	.29	.29	.36	.20	.30	1	.50	.50	.50	.27	.27	.25	.25	.18	.07	.05	.06	.05	.05	.02
19. Sizeable	0	0	0	0	0	0	.01	.01	.02	.02	.02	.11	.16	.16	.21	.12	.18	.50	1	1	.84	.47	.47	.43	.42	.32	.09	.07	.08	.08	.07	.03
20. Quite a bit	0	0	0	0	0	0	.01	.01	.02	.02	.02	.11	.16	.16	.21	.12	.18	.50	1	1	.84	.47	.47	.43	.42	.32	.09	.07	.08	.08	.07	.03
21. Considerable amount	0	0	0	0	0	0	.01	.01	.02	.02	.02	.11	.16	.15	.20	.11	.16	.50	.84	.84	1	.49	.49	.44	.45	.32	.09	.08	.08	.08	.08	.03
22. Substantial amount	0	0	0	0	0	0	0	.01	.01	.01	.01	.06	.09	.08	.11	.06	.09	.27	.47	.47	.49	1	.98	.82	.79	.63	.15	.13	.14	.14	.13	.05
23. A lot	0	0	0	0	0	0	0	.01	.01	.01	.01	.06	.09	.08	.11	.06	.09	.27	.47	.47	.49	.98	1	.83	.79	.63	.15	.13	.14	.13	.13	.05
24. High amount	0	0	0	0	0	0	0	.01	.01	.01	.01	.06	.08	.07	.10	.05	.08	.25	.43	.43	.44	.82	.83	1	.89	.70	.17	.14	.16	.15	.14	.06
25. Very sizeable	0	0	0	0	0	0	0	.01	.01	.01	.01	.06	.08	.07	.10	.05	.08	.25	.42	.42	.45	.79	.79	.89	1	.64	.15	.14	.14	.13	.13	.05
26. Large	0	0	0	0	0	0	0	0	0	0	0	.04	.06	.05	.07	.03	.05	.18	.32	.32	.32	.63	.63	.70	.64	1	.17	.15	.16	.15	.15	.05
27. Very large	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.01	.02	.01	.01	.07	.09	.09	.09	.15	.15	.17	.15	.17	1	.67	.86	.70	.68	.21
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	.01	.02	.01	.02	.01	.01	.05	.07	.07	.08	.13	.13	.14	.14	.15	.67	1	.66	.68	.68	.22
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.01	.02	.01	.01	.06	.08	.08	.08	.14	.14	.16	.14	.16	.86	.66	1	.83	.80	.25
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	.01	.02	.01	.02	.01	.01	.05	.08	.08	.08	.14	.13	.15	.13	.15	.70	.68	.83	1	.96	.25
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	.01	.02	.01	.02	.01	.01	.05	.07	.07	.08	.13	.13	.14	.13	.15	.68	.68	.80	.96	1	.26
32. Maximum amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	0	.01	0	0	.02	.03	.03	.03	.05	.05	.06	.05	.05	.21	.22	.25	.25	.26	1

Table 7
Similarity matrix when the Jaccard similarity measure is used.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. None to very little	1	.80	.77	.75	.64	.65	.11	.11	.16	.13	.08	.05	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. Teeny-weeny	.80	1	.63	.61	.51	.51	.12	.12	.17	.14	.08	.05	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. A smidgen	.77	.63	1	.97	.80	.82	.19	.18	.24	.21	.14	.09	.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.75	.61	.97	1	.81	.84	.20	.19	.24	.21	.14	.09	.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very small	.64	.51	.80	.81	1	.92	.18	.17	.23	.19	.13	.08	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very little	.65	.51	.82	.84	.92	1	.20	.19	.25	.21	.14	.09	.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.11	.12	.19	.20	.18	.20	1	.62	.51	.46	.40	.21	.11	.02	.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8. Little	.11	.12	.18	.19	.17	.19	.62	1	.85	.77	.66	.35	.22	.10	.12	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9. Low amount	.16	.17	.24	.24	.23	.25	.51	.85	1	.83	.65	.35	.21	.10	.12	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10. Small	.13	.14	.21	.21	.19	.21	.46	.77	.83	1	.74	.39	.24	.11	.13	.04	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11. Somewhat small	.08	.08	.14	.14	.13	.14	.40	.66	.65	.74	1	.43	.26	.12	.13	.03	.03	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12. Some	.05	.05	.09	.09	.08	.09	.21	.35	.35	.39	.43	1	.71	.56	.54	.37	.38	.26	.16	.16	.16	.08	.08	.08	.08	.05	0	0	0	0	0	0
13. Some to moderate	.01	.01	.04	.04	.03	.04	.11	.22	.21	.24	.26	.71	1	.75	.70	.45	.51	.33	.19	.19	.19	.10	.10	.09	.10	.06	0	0	0	0	0	0
14. Moderate amount	0	0	0	0	0	0	.02	.10	.10	.11	.12	.56	.75	1	.79	.60	.63	.37	.21	.21	.21	.10	.10	.10	.10	.06	0	0	0	0	0	0
15. Fair amount	0	0	0	0	0	0	.04	.12	.12	.13	.13	.54	.70	.79	1	.52	.69	.42	.25	.25	.25	.12	.12	.12	.12	.08	0	0	0	0	0	0
16. Medium	0	0	0	0	0	0	0	.03	.03	.04	.03	.37	.45	.60	.52	1	.76	.37	.19	.19	.19	.07	.07	.07	.07	.03	0	0	0	0	0	0
17. Modest amount	0	0	0	0	0	0	0	.03	.03	.03	.03	.38	.51	.63	.69	.76	1	.46	.26	.26	.25	.11	.11	.11	.11	.07	0	0	0	0	0	0
18. Good amount	0	0	0	0	0	0	0	.03	.03	.03	.02	.26	.33	.37	.42	.37	.46	1	.64	.64	.63	.40	.39	.38	.39	.32	.10	.10	.10	.10	.10	.03
19. Sizeable	0	0	0	0	0	0	0	0	0	0	0	.16	.19	.21	.25	.19	.26	.64	1	1	.90	.52	.52	.51	.50	.43	.11	.12	.11	.11	.11	.02
20. Quite a bit	0	0	0	0	0	0	0	0	0	0	0	.16	.19	.21	.25	.19	.26	.64	1	1	.90	.52	.52	.51	.50	.43	.11	.12	.11	.11	.11	.02
21. Considerable amount	0	0	0	0	0	0	0	0	0	0	0	.16	.19	.21	.25	.19	.25	.63	.90	.90	1	.60	.60	.58	.58	.50	.14	.15	.14	.14	.14	.04
22. Substantial amount	0	0	0	0	0	0	0	0	0	0	0	.08	.10	.10	.12	.07	.11	.40	.52	.52	.60	1	.99	.95	.88	.73	.22	.23	.22	.22	.22	.08
23. A lot	0	0	0	0	0	0	0	0	0	0	0	.08	.10	.10	.12	.07	.11	.39	.52	.52	.60	.99	1	.94	.87	.72	.22	.23	.22	.22	.22	.08
24. High amount	0	0	0	0	0	0	0	0	0	0	0	.08	.09	.10	.12	.07	.11	.38	.51	.51	.58	.95	.94	1	.90	.77	.22	.22	.21	.21	.21	.07
25. Very sizeable	0	0	0	0	0	0	0	0	0	0	0	.08	.10	.10	.12	.07	.11	.39	.50	.50	.58	.88	.87	.90	1	.72	.25	.24	.24	.24	.23	.08
26. Large	0	0	0	0	0	0	0	0	0	0	0	.05	.06	.06	.08	.03	.07	.32	.43	.43	.50	.73	.72	.77	.72	1	.21	.20	.19	.20	.19	.05
27. Very large	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.10	.11	.11	.14	.22	.22	.22	.25	.21	1	.67	.91	.79	.76	.40
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.10	.12	.12	.15	.23	.23	.22	.24	.20	.67	1	.74	.85	.88	.52
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.10	.11	.11	.14	.22	.22	.21	.24	.19	.91	.74	1	.87	.84	.44
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.10	.11	.11	.14	.22	.22	.21	.24	.20	.79	.85	.87	1	.97	.50
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.10	.11	.11	.14	.22	.22	.21	.23	.19	.76	.88	.84	.97	1	.52
32. Maximum amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.03	.02	.02	.04	.08	.08	.07	.08	.05	.40	.52	.44	.50	.52	1

Table 8
Groups of similar words when the Jaccard similarity measure is used. All words to the left of or in the column of s_J^*, i.e., $s_J \geqslant s_J^*$, are similar to a (numbered) word at a similarity value that is *at least* s_J^*.

Word	$s_J \geqslant 0.9$	$s_J \geqslant 0s.8$	$s_J \geqslant 0.7$	$s_J \geqslant 0.6$	$s_J \geqslant 0.5$
1. None to very little			Teeny-weeny A smidgen Tiny	Very little Very small	
2. Teeny-weeny			None to very little	A smidgen Tiny	Very little Very small
3. A smidgen	Tiny	Very little	Very small None to very little	Teeny-weeny	
4. Tiny	A smidgen	Very little Very small	None to very little	Teeny-weeny	
5. Very small	Very little	Tiny	A smidgen	None to very little	Teeny-weeny
6. Very little	Very small	Tiny A smidgen		None to very little	Teeny-weeny
7. A bit				Little	Low amount
8. Little		Low amount	Small	Somewhat small A bit	
9. Low amount		Little Small		Somewhat small	A bit
10. Small		Low amount	Little Somewhat small		
11. Somewhat small			Small	Little Low amount	
12. Some			Some to moderate		Moderate amount Fair amount
13. Some to moderate			Moderate amount Some	Fair amount	Modest amount
14. Moderate amount			Fair amount Some to moderate	Modest amount	Medium Some
15. Fair amount			Moderate amount	Some to moderate Modest amount	Some Medium
16. Medium			Modest amount		Moderate amount Fair amount
17. Modest amount			Medium	Fair amount Moderate amount	Some to moderate
18. Good amount				Sizeable Quite a bit Considerable amount	
19. Sizeable		Considerable amount		Good amount	A lot Substantial amount High amount Very sizeable
20. Quite a bit		Considerable amount		Good amount	A lot Substantial amount High amount Very sizeable
21. Considerable amount		Sizeable Quite a bit		Good amount	Substantial amount A lot High amount Very sizeable
22. Substantial amount	A lot High amount	Very sizeable	Large		Considerable amount Sizeable Quite a bit
23. A lot	Substantial amount High amount	Very sizeable	Large		Considerable amount Sizeable Quite a bit
24. High amount	Substantial amount A lot Very sizeable		Large		Considerable amount Sizeable Quite a bit
25. Very sizeable	High amount	Substantial amount A lot	Large		Considerable amount Sizeable Quite a bit
26. Large			High amount Substantial amount Very sizeable A lot		
27. Very large	Huge amount		Very high amount Extreme amount	Humongous amount	
28. Humongous amount		Extreme amount Very high amount	Huge amount	Very large	Maximum amount

Tabel 8 (*continued*)

Word	$s_J \geqslant 0.9$	$s_J \geqslant 0s.8$	$s_J \geqslant 0.7$	$s_J \geqslant 0.6$	$s_J \geqslant 0.5$
29. Huge amount	Very large	Very high amount Extreme amount	Humongous amount		
30. Very high amount	Extreme amount	Huge amount Humongous amount	Very large		Maximum amount
31. Extreme amount	Very high amount	Humongous amount Huge amount	Very large		Maximum amount
32. Maximum amount					Extreme amount Humongous amount Very high amount

To begin, we review how cardinality, fuzziness, variance and skewness can be computed for an IT2 FS. In Sections 5.1–5.4 results are stated without proofs, because the latter can be found in [34].

5.1. Cardinality of an IT2 FS

Szmidt and Kacprzyk [27] derived an interval cardinality for intuitionistic fuzzy sets (IFS) [1]. Though IFSs are different from IT2 FSs, Atanassov and Gargov [1] showed that every IFS can be mapped to an interval valued FS, which is an IT2 FS under a different name. Using Atanassov and Gargov's mapping, Szmidt and Kacprzyk's interval cardinality for an IT2 FS $\widetilde{A}$ is

$$P_{SK}(\widetilde{A}) = [p_{DT}(LMF(\widetilde{A})), p_{DT}(UMF(\widetilde{A}))] \tag{33}$$

where $p_{DT}(A)$ is De Luca and Termini's [4] definition of T1 FS cardinality, i.e.,

$$p_{DT}(A) = \int_X \mu_A(x)dx. \tag{34}$$

A *normalized cardinality* for a T1 FS is used in this paper, and it is defined by discretizing $p_{DT}(A)$, i.e.,

$$p(A) = \frac{|X|}{N}\sum_{i=1}^{N}\mu_A(x_i). \tag{35}$$

where $|X| = x_N - x_1$ is the length of the universe of discourse used in the computation.

Definition 4. The cardinality of an IT2 FS $\widetilde{A}$ is the union of all cardinalities of its embedded T1 FSs A_e, i.e.,

$$P(\widetilde{A}) \equiv \bigcup_{\forall A_e} p(A_e) = [p_l(\widetilde{A}), p_r(\widetilde{A})], \tag{36}$$

where

$$p_l(\widetilde{A}) = \min_{\forall A_e} p(A_e) \tag{37}$$

$$p_r(\widetilde{A}) = \max_{\forall A_e} p(A_e). \tag{38}$$

Theorem 3. *$p_l(\widetilde{A})$ and $p_r(\widetilde{A})$ in (37) and (38) can be computed as*

$$p_l(\widetilde{A}) = p(LMF(\widetilde{A})) \tag{39}$$

$$p_r(\widetilde{A}) = p(UMF(\widetilde{A})). \tag{40}$$

Observe that $P(\widetilde{A})$ is very similar to $P_{SK}(\widetilde{A})$, except that a different T1 FS cardinality definition is used.

Another useful concept is the *average cardinality* of $\widetilde{A}$, which is defined as the average of its minimum and maximum cardinalities, i.e.,

$$p(\widetilde{A}) = \frac{p(LMF(\widetilde{A})) + p(UMF(\widetilde{A}))}{2}. \tag{41}$$

$p(\widetilde{A})$ has been used in Section 4 to define the VSM and Jaccard's similarity measure.

5.2. Fuzziness (entropy) of an IT2 FS

The fuzziness (entropy) of an IT2 FS quantifies the amount of vagueness in it.

Definition 5. The fuzziness $F(\widetilde{A})$ of an IT2 FS $\widetilde{A}$ is the union of the fuzziness of all its embedded T1 FSs A_e, i.e.,

$$F(\widetilde{A}) \equiv \bigcup_{\forall A_e} f(A_e) = [f_l(\widetilde{A}), f_r(\widetilde{A})], \tag{42}$$

where $f_l(\widetilde{A})$ and $f_r(\widetilde{A})$ are the minimum and maximum of the fuzziness of all A_e, respectively, i.e.

$$f_l(\widetilde{A}) = \min_{\forall A_e} f(A_e) \tag{43}$$

$$f_r(\widetilde{A}) = \max_{\forall A_e} f(A_e). \tag{44}$$

Theorem 4. *Let $f(A_e)$ be Yager's fuzziness measure [40]:*

$$f(A_e) = 1 - \frac{1}{N}\sum_{i=1}^{N} |2\mu_A(x_i) - 1)|, \tag{45}$$

Additionally, let A_{e1} be defined as

$$\mu_{A_{e1}}(x) = \begin{cases} \bar{\mu}_{\widetilde{A}}(x), & \bar{\mu}_{\widetilde{A}}(x) \text{ is further away from } 0.5 \text{ than } \underline{\mu}_{\widetilde{A}}(x) \\ \underline{\mu}_{\widetilde{A}}(x), & \text{otherwise} \end{cases} \tag{46}$$

and A_{e2} be defined as

$$\mu_{A_{e2}}(x) = \begin{cases} \bar{\mu}_{\widetilde{A}}(x), & \text{both } \bar{\mu}_{\widetilde{A}}(x) \text{ and } \underline{\mu}_{\widetilde{A}}(x) \text{ arebelow } 0.5 \\ \underline{\mu}_{\widetilde{A}}(x), & \text{both } \bar{\mu}_{\widetilde{A}}(x) \text{ and } \underline{\mu}_{\widetilde{A}}(x) \text{ areabove } 0.5 \\ 0.5, & \text{otherwise} \end{cases} \tag{47}$$

Then (43) and (44) can be computed as

$$f_l(\widetilde{A}) = f(A_{e1}) \tag{48}$$

$$f_r(\widetilde{A}) = f(A_{e2}). \tag{49}$$

5.3. Variance of an IT2 FS

The variance of a T1 FS A measures its compactness, i.e. a smaller (larger) variance means A is more (less) compact.

Definition 6. The relative variance of an embedded T1 FS A_e to an IT2 FS $\widetilde{A}$, $v_{\widetilde{A}}(A_e)$, is defined as

$$v_{\widetilde{A}}(A_e) = \frac{\sum_{i=1}^{N}[x_i - c(\widetilde{A})]^2 \mu_{A_e}(x_i)}{\sum_{i=1}^{N}\mu_{A_e}(x_i)}, \tag{50}$$

where $c(\widetilde{A})$ is the average centroid of $\widetilde{A}$ (see (9)).

Definition 7. The variance of an IT2 FS $\widetilde{A}$, $V(\widetilde{A})$, is the union of relative variance of all its embedded T1 FSs A_e, i.e.,

$$V(\widetilde{A}) \equiv \bigcup_{\forall A_e} v_{\widetilde{A}}(A_e) = [v_l(\widetilde{A}), v_r(\widetilde{A})], \tag{51}$$

where $v_l(\widetilde{A})$ and $v_r(\widetilde{A})$ are the minimum and maximum relative variance of all A_e, respectively, i.e.

$$v_l(\widetilde{A}) = \min_{\forall A_e} v_{\widetilde{A}}(A_e) \tag{52}$$

$$v_r(\widetilde{A}) = \max_{\forall A_e} v_{\widetilde{A}}(A_e). \tag{53}$$

$v_l(\widetilde{A})$ and $v_r(\widetilde{A})$ can be computed by KM algorithms.

5.4. Skewness of an IT2 FS

The *skewness* of a T1 FS A, $s(A)$, is an indicator of its symmetry. $s(A)$ is smaller than zero when A skews to the right, is larger than zero when A skews to the left, and is equal to zero when A is symmetrical.

Table 9
Uncertainty measures for the 32 word FOUs.

Word	Area of FOU	$C(\widetilde{A})$	$P(\widetilde{A})$	$F(\widetilde{A})$	$V(\widetilde{A})$	$S(\widetilde{A})$
1. None to very little	0.70	[0.22, 0.73]	[0.35, 1.05]	[0.06, 0.66]	[0.06, 0.38]	[−0.03, 0.31]
2. Teeny-weeny	0.98	[0.05, 1.07]	[0.07, 1.05]	[0, 0.74]	[0.06, 0.74]	[−0.14, 0.61]
3. A smidgen	1.11	[0.21, 1.05]	[0.33, 1.44]	[0.02, 0.70]	[0.10, 0.83]	[−0.10, 0.95]
4. Tiny	1.16	[0.21, 1.06]	[0.33, 1.49]	[0.01, 0.71]	[0.10, 0.85]	[−0.10, 0.96]
5. Very small	0.92	[0.39, 0.93]	[0.62, 1.55]	[0.04, 0.67]	[0.11, 0.52]	[−0.07, 0.47]
6. Very little	1.09	[0.33, 1.01]	[0.53, 1.63]	[0.02, 0.69]	[0.11, 0.68]	[−0.08, 0.71]
7. A bit	1.13	[1.42, 2.08]	[0.53, 1.66]	[0.09, 0.75]	[0.09, 0.52]	[−0.16, 0.43]
8. Little	2.32	[1.31, 2.95]	[0.30, 2.62]	[0.02, 0.81]	[0.10, 1.73]	[−1.03, 2.77]
9. Low amount	2.81	[0.92, 3.46]	[0.08, 2.89]	[0, 0.82]	[0.02, 2.63]	[−3.30, 4.58]
10. Small	2.81	[1.29, 3.34]	[0.20, 3.01]	[0, 0.83]	[0.03, 2.06]	[−2.66, 2.85]
11. Somewhat small	2.34	[1.76, 3.43]	[0.19, 2.53]	[0, 0.83]	[0.03, 1.43]	[−1.80, 1.35]
12. Some	4.74	[2.04, 5.77]	[0.23, 4.97]	[0, 0.83]	[0.08, 6.29]	[−12.43, 16.59]
13. Some to moderate	4.07	[3.02, 6.11]	[0.26, 4.33]	[0, 0.82]	[0.05, 4.58]	[−9.72, 8.47]
14. Moderate amount	3.09	[3.74, 6.16]	[0.17, 3.26]	[0, 0.82]	[0.03, 2.74]	[−3.55, 4.83]
15. Fair amount	3.45	[3.85, 6.41]	[0.25, 3.70]	[0, 0.82]	[0.06, 3.25]	[−6.13, 4.59]
16. Medium	2.00	[4.19, 6.19]	[0.04, 2.03]	[0, 0.80]	[0.01, 1.52]	[−1.56, 1.91]
17. Modest amount	2.34	[4.57, 6.24]	[0.19, 2.53]	[0, 0.83]	[0.03, 1.43]	[−1.35, 1.80]
18. Good amount	3.83	[5.11, 7.89]	[0.29, 4.12]	[0, 0.83]	[0.05, 3.85]	[−7.03, 7.03]
19. Sizeable	2.92	[6.17, 8.15]	[0.35, 3.26]	[0, 0.82]	[0.10, 2.30]	[−4.00, 2.21]
20. Quite a bit	2.92	[6.17, 8.15]	[0.35, 3.26]	[0, 0.82]	[0.10, 2.30]	[−4.00, 2.21]
21. Considerable amount	3.31	[5.97, 8.52]	[0.19, 3.49]	[0, 0.83]	[0.08, 3.09]	[−6.07, 3.62]
22. Substantial amount	2.61	[6.95, 8.86]	[0.23, 2.84]	[0, 0.82]	[0.08, 2.01]	[−3.36, 1.61]
23. A lot	2.59	[6.99, 8.83]	[0.26, 2.85]	[0, 0.82]	[0.07, 1.92]	[−3.16, 1.55]
24. High amount	2.46	[7.19, 8.82]	[0.38, 2.84]	[0.02, 0.82]	[0.13, 1.83]	[−3.00, 1.08]
25. Very sizeable	2.79	[6.95, 9.10]	[0.17, 2.96]	[0, 0.83]	[0.13, 2.50]	[−4.49, 1.69]
26. Large	1.87	[7.50, 8.75]	[0.32, 2.19]	[0.04, 0.80]	[0.10, 1.18]	[−1.55, 0.47]
27. Very large	0.92	[9.03, 9.57]	[0.68, 1.60]	[0.06, 0.66]	[0.12, 0.57]	[−0.55, 0.08]
28. Humongous amount	1.27	[8.70, 9.91]	[0.13, 1.40]	[0, 0.73]	[0.10, 1.18]	[−1.33, 0.23]
29. Huge amount	0.96	[9.03, 9.65]	[0.55, 1.51]	[0.05, 0.67]	[0.11, 0.63]	[−0.66, 0.08]
30. Very high amount	1.09	[8.96, 9.78]	[0.35, 1.44]	[0.02, 0.70]	[0.10, 0.82]	[−0.92, 0.09]
31. Extreme amount	1.07	[8.96, 9.79]	[0.33, 1.40]	[0.03, 0.69]	[0.10, 0.83]	[−0.94, 0.09]
32. Maximum amount	0.50	[9.50, 9.87]	[0.21, 0.70]	[0.04, 0.67]	[0.03, 0.18]	[−0.10, 0.01]

Definition 8. The relative skewness of an embedded T1 FS A_e to an IT2 FS $\widetilde{A}$, $s_{\widetilde{A}}(A_e)$, is defined as

$$s_{\widetilde{A}}(A_e) = \frac{\sum_{i=1}^{N}[x_i - c(\widetilde{A})]^3 \mu_{A_e}(x_i)}{\sum_{i=1}^{N}\mu_{A_e}(x_i)}, \tag{54}$$

where $c(\widetilde{A})$ is the average centroid of $\widetilde{A}$ (see (9)).

Definition 9. The skewness of an IT2 FS $\widetilde{A}$, $S(\widetilde{A})$, is the union of relative skewness of all its embedded T1 FSs A_e, i.e.,

$$S(\widetilde{A}) \equiv \bigcup_{\forall A_e} s_{\widetilde{A}}(A_e) = [s_l(\widetilde{A}), s_r(\widetilde{A})], \tag{55}$$

where $s_l(\widetilde{A})$ and $s_r(\widetilde{A})$ are the minimum and maximum relative skewness of all A_e, respectively, i.e.

$$s_l(\widetilde{A}) = \min_{\forall A_e} s_{\widetilde{A}}(A_e) \tag{56}$$

$$s_r(\widetilde{A}) = \max_{\forall A_e} s_{\widetilde{A}}(A_e). \tag{57}$$

$s_l(\widetilde{A})$ and $s_r(\widetilde{A})$ can be computed by KM algorithms.

5.5. Comparative studies

The areas of the 32 word FOUs, as well as the five uncertainty measures for them, are summarized in Table 9. Clearly, it is difficult to know what to do with all these measures. In this section, we study whether or not all are needed.

Average cardinality, $p(\widetilde{A})$, has been defined in (41). Additionally, we introduce the following quantities that are functions of our uncertainty measures[11]:

[11] In probability theory, the mean of a random variable is not an uncertainty measure. Analogously, we may view the average centroid $c(\widetilde{A})$ as the "mean" of $\widetilde{A}$, which indicates whether $\widetilde{A}$ is "large" or "small" but is not an uncertainty measure.

Table 10
Correlations among different uncertainty measures.

Area	Area	Intra-personal				Inter-personal				
		$p(\tilde{A})$	$f(\tilde{A})$	$v(\tilde{A})$	$\lvert s(\tilde{A})\rvert$	$\delta_c(\tilde{A})$	$\delta_p(\tilde{A})$	$\delta_f(\tilde{A})$	$\delta_v(\tilde{A})$	$\delta_s(\tilde{A})$
	1	.99	.91	.98	.88	1	1	.93	.97	.88
$p(\tilde{A})$	.99	1	.95	.95	.84	.97	.99	.96	.94	.84
$f(\tilde{A})$	.91	.95	1	.84	.67	.90	.91	1	.81	.67
$v(\tilde{A})$	.98	.95	.84	1	.96	.98	.98	.86	1	.96
$s(\tilde{A})$	.88	.84	.67	.96	1	.89	.88	.69	.97	1
$\delta_c(\tilde{A})$	1	.97	.90	.98	.89	1	1	.92	.97	.89
$\delta_p(\tilde{A})$	1	.99	.91	.98	.88	1	1	.93	.97	.88
$\delta_f(\tilde{A})$	.93	.96	1	.86	.69	.92	.93	1	.83	.69
$\delta_v(\tilde{A})$	.97	.94	.81	1	.97	.97	.97	.83	1	.97
$\delta_s(\tilde{A})$	.88	.84	.67	.96	1	.89	.88	.69	.97	1

$$f(\tilde{A}) \equiv \frac{f_r(\tilde{A}) + f_l(\tilde{A})}{2} \tag{58}$$

$$v(\tilde{A}) \equiv \frac{v_r(\tilde{A}) + v_l(\tilde{A})}{2} \tag{59}$$

$$|s(\tilde{A})| \equiv \frac{|s_r(\tilde{A})| + |s_l(\tilde{A})|}{2} \tag{60}$$

$$\delta_c(\tilde{A}) \equiv c_r(\tilde{A}) - c_l(\tilde{A}) \tag{61}$$

$$\delta_p(\tilde{A}) \equiv p_r(\tilde{A}) - p_l(\tilde{A}) \tag{62}$$

$$\delta_f(\tilde{A}) \equiv f_r(\tilde{A}) - f_l(\tilde{A}) \tag{63}$$

$$\delta_v(\tilde{A}) \equiv v_r(\tilde{A}) - v_l(\tilde{A}) \tag{64}$$

$$\delta_s(\tilde{A}) \equiv s_r(\tilde{A}) - s_l(\tilde{A}) \tag{65}$$

Observe that:

(i) $p(\tilde{A}), f(\tilde{A})$, $v(\tilde{A})$ and $|s(\tilde{A})|$ are intra-personal uncertainty measures[12], because they measure the average uncertainties of the embedded T1 FSs; and
(ii) $\delta_c(\tilde{A})$, $\delta_p(\tilde{A})$, $\delta_f(\tilde{A})$, $\delta_v(\tilde{A})$ and $\delta_s(\tilde{A})$ are inter-personal uncertainty measures, because they indicate how the embedded T1 FSs are different from each other.

The correlation between any two of these nine quantities (called q_1 and q_2) is computed as

$$correlation(q_1, q_2) = \frac{\sum_{i=1}^{32} q_1(\tilde{A}_i) q_2(\tilde{A}_i)}{\sqrt{[\sum_{i=1}^{32} q_1^2(\tilde{A}_i)][\sum_{j=1}^{32} q_2^2(\tilde{A}_j)]}} \tag{66}$$

and all correlations are summarized in Table 10, along with the areas of the FOUs. Observe that:

(i) All nine quantities have strong correlation with the area of the FOU (see the area row and column). This is because as the area of the FOU increases, both intra-personal and inter-personal uncertainties increase.
(ii) Among the four intra-personal uncertainty measures (see the 4×4 matrix in the intra-personal sub-table), average cardinality $p(\tilde{A})$ and average variance $v(\tilde{A})$ have the strongest correlation with all other intra-personal uncertainty measures; hence, they are the most representative[13] intra-personal uncertainty measures.
(iii) Among the five inter-personal uncertainty measures (see the 5×5 matrix in the inter-personal sub-table), $\delta_c(\tilde{A})$ and $\delta_p(\tilde{A})$ have correlation 1, and both of them have the strongest correlation with all other inter-personal uncertainty measures; hence, they are the most representative inter-personal uncertainty measures.

In summary, *cardinality* is the most important uncertainty measure for an IT2 FS: its center is a representative intra-personal uncertainty measure, and its length is a representative inter-personal uncertainty measure.

[12] Any value within the interval $[p_l(\tilde{A}), p_r(\tilde{A})]$ is an intra-personal uncertainty measure because it corresponds to the cardinality of an embedded T1 FS, i.e., a single person's opinion; however, $p(\tilde{A})$ is used because it is the most representative one. The other three quantities can be understood in a similar way.
[13] By *representative* we mean that when $p(\tilde{A})$ or $v(\tilde{A})$ is large, we have high confidence that the other three intra-personal uncertainty measures are also large; hence, only $p(\tilde{A})$ or $v(\tilde{A})$ needs to be computed for intra-personal uncertainty.

Because the length of the centroid is a representative inter-personal uncertainty measure, and the average centroid can be used in ranking IT2 FSs, the centroid is also a very important characteristic of IT2 FSs.

6. Conclusions

In this paper, several ranking methods, similarity measures and uncertainty measures for IT2 FSs have been evaluated using real survey data. It has been shown that:

(i) Our new centroid-based ranking method is better than Mitchell's ranking method for IT2 FSs.
(ii) The Jaccard similarity measure is better than all other similarity measures for IT2 FSs.
(iii) Cardinality is the most representative uncertainty measure for an IT2 FS: its center is a representative intra-personal uncertainty measure, and its length is a representative inter-personal uncertainty measure.
(iv) Centroid is a very important characteristic for IT2 FSs: its center can be used in ranking, and its length is a representative inter-personal uncertainty measure.

These results, which can easily be re-done for new data sets that a reader collects, should help people better understand the uncertainties associated with linguistic terms and hence how to use the uncertainties effectively in survey design and linguistic information processing.

Acknowledgements

This work was supported by the 2007 IEEE Computational Intelligence Society Walter Karplus Summer Research Grant. The authors would like to thank Professor David V. Budescu, University of Illinois at Urbana-Champaign, for his very helpful comments.

References

[1] K. Atanassov, G. Gargov, Interval valued intuitionistic fuzzy sets, Fuzzy Sets and Systems 31 (1989) 343–349.
[2] J.J. Buckley, T. Feuring, Computing with words in control, in: L.A. Zadeh, J. Kacprzyk (Eds.), Computing with Words in Information/Intelligent Systems 2: Applications, Physica-Verlag, Heidelberg, 1999, pp. 289–304.
[3] 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.
[4] A. De Luca, S. Termini, A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, Information and Computation 20 (1972) 301–312.
[5] M.B. Gorzalczany, A method of inference in approximate reasoning based on interval-valued fuzzy sets, Fuzzy Sets and Systems 21 (1987) 1–17.
[6] D. Harmanec, Measures of uncertainty and information, <http://www.sipta.org/documentation/summary_measures/main.html>, 1999.
[7] P. Jaccard, Nouvelles recherches sur la distribution florale, Bulletin de la Societe de Vaud des Sciences Naturelles 44 (1908) 223.
[8] N.N. Karnik, J.M. Mendel, Centroid of a type-2 fuzzy set, Information Sciences 132 (2001) 195–220.
[9] G.J. Klir, Principles of uncertainty: what are they? why do we need them?, Fuzzy Sets and Systems 74 (1995) 15–31
[10] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, Upper Saddle River, NJ, 1995.
[11] F. Liu, J.M. Mendel, Encoding words into interval type-2 fuzzy sets using an interval approach, IEEE Transactions on Fuzzy Systems 16 (6) (2008) 1503–1521.
[12] M. Margaliot, G. Langholz, Fuzzy control of a benchmark problem: a computing with wordsapproach, in: Joint 9th IFSA World Congress and 20th NAFIPS International Conference, vol. 5, Vancouver, Canada, 2001, pp. 3065–3069.
[13] J.M. Mendel, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice-Hall, Upper Saddle River, NJ, 2001.
[14] J.M. Mendel, Computing with words, when words can mean different things to different people, in: Proceedings third International ICSC Symposium on Fuzzy Logic and Applications, Rochester, NY, 1999, pp. 158–164.
[15] J.M. Mendel, The perceptual computer: an architecture for computing with words, in: Proceedings of the FUZZ-IEEE, Melbourne, Australia, 2001, pp. 35–38.
[16] J.M. Mendel, An architecture for making judgments using computing with words, International Journal of Applied Mathematics and Computer Science 12 (3) (2002) 325–335.
[17] J.M. Mendel, Computing with words and its relationships with fuzzistics, Information Sciences 177 (2007) 988–1006.
[18] J.M. Mendel, D. Wu, Perceptual reasoning: a new computing with words engine, in: Proceedings of the IEEE International Conference on Granular Computing, Silicon Valley, CA, 2007, pp. 446–451.
[19] J.M. Mendel, D. Wu, Perceptual reasoning for perceptual computing, IEEE Transactions on Fuzzy systems 16 (6) (2008) 1550–1564.
[20] 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.
[21] J.M. Mendel, H. Wu, New results about the centroid of an interval type-2 fuzzy set,including the centroid of a fuzzy granule, Information Sciences 177 (2007) 360–377.
[22] J.M. Mendel, H. Wu, Type-2 fuzzistics for non-symmetric interval type-2 fuzzy sets: forward problems, IEEE Transactions on Fuzzy Systems 15 (5) (2007) 916–930.
[23] H.B. Mitchell, Pattern recognition using type-II fuzzy sets, Information Sciences 170 (2-4) (2005) 409–418.
[24] H.B. Mitchell, Ranking type-2 fuzzy numbers, IEEE Transactions on Fuzzy Systems 14 (2) (2006) 287–294.
[25] V. Novák, Mathematical fuzzy logic in modeling of natural language semantics, in: P. Wang, D. Ruan, E. Kerre (Eds.), Fuzzy Logic – A Spectrum of Theoretical and Practical Issues, Elsevier, Berlin, 2007, pp. 145–182.
[26] K.S. Schmucker, Fuzzy Sets, Natural Language Computations, and Risk Analysis, Computer Science Press, Rockville, MD, 1984.
[27] E. Szmidt, J. Kacprzyk, Entropy for intuitionistic fuzzy sets, Fuzzy Sets and Systems 118 (2001) 467–477.
[28] R.M. Tong, P.P. Bonissone, A linguistic approach to decision making with fuzzy sets, IEEE Transactions on Systems, Man, and Cybernetics 10 (1980) 716–723.
[29] 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.
[30] X. Wang, E.E. Kerre, Reasonable properties for the ordering of fuzzy quantities (I), Fuzzy Sets and Systems 118 (2001) 375–387.

[31] X. Wang, E.E. Kerre, Reasonable properties for the ordering of fuzzy quantities (II), Fuzzy Sets and Systems 118 (2001) 387–405.
[32] D. Wu, J.M. Mendel, The linguistic weighted average, in: Proceedings of the FUZZ-IEEE, Vancouver, BC, Canada, 2006, pp. 566–573.
[33] D. Wu, J.M. Mendel, Aggregation using the linguistic weighted average and interval type-2 fuzzy sets, IEEE Transactions on Fuzzy Systems 15 (6) (2007) 1145–1161.
[34] D. Wu, J.M. Mendel, Uncertainty measures for interval type-2 fuzzy sets, Information Sciences 177 (23) (2007) 5378–5393.
[35] D. Wu, J.M. Mendel, Corrections to "Aggregation using the linguistic weighted average and interval type-2 fuzzy sets, IEEE Transactions on Fuzzy Systems 16 (6) (2008) 1664–1666.
[36] D. Wu, J.M. Mendel, Enhanced Karnik–Mendel Algorithms, IEEE Transactions on Fuzzy Systems, in press.
[37] D. Wu, J.M. Mendel, A vector similarity measure for linguistic approximation: interval type-2 and type-1 fuzzy sets, Information Sciences 178 (2) (2008) 381–402.
[38] R. Yager, Approximate reasoning as a basis for computing with words, in: L.A. Zadeh, J. Kacprzyk (Eds.), Computing with Words in Information/ Intelligent Systems 1: Foundations, Physica-Verlag, Heidelberg, 1999, pp. 50–77.
[39] R.R. Yager, Ranking fuzzy subsets over the unit interval, in: Proceedings of the IEEE Conference on Decision and Control, vol. 17, 1978, pp. 1435–1437.
[40] R.R. Yager, A measurement-informational discussion of fuzzy union and fuzzy intersection, International Journal of Man–Machine Studies 11 (1979) 189–200.
[41] R.R. Yager, On the retranslation process in zadeh's paradigm of computing with words, IEEE Transactions on Systems, Man and Cybernetics, Part B 34 (2) (2004) 1184–1195.
[42] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning-1, Information Sciences 8 (1975) 199–249.
[43] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Transactions on Fuzzy Systems 4 (1996) 103–111.
[44] 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.
[45] W. Zeng, H. Li, Relationship between similarity measure and entropy of interval valued fuzzy sets, Fuzzy Sets and Systems 157 (2006) 1477–1484.