A Comparative Study of Ranking Methods, Similarity Measures and Uncertainty Measures for Interval Type-2 Fuzzy Sets

Dongrui Wu, *Student Member (#80365605), IEEE*
Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA
Email: dongruiw@usc.edu

Abstract

Ranking methods, similarity measures and uncertainty measures are very important concepts for interval type-2 fuzzy sets (IT2 FSs). For example, a ranking method can be used to find the largest (best) IT2 FSs in a vocabulary or after aggregation, a similarity measure can be used to map the output of a computing with words engine to a word in the vocabulary, and, uncertainty measures can be used in system design using basic uncertainty principles. So far, there is only one ranking method for IT2 FSs, whereas there are many similarity and uncertainty measures (which causes confusion). The objectives of this report are to evaluate ranking methods, similarity measures and uncertainty measures for IT2 FSs based on real survey data, test ranking methods and the LWA against psychological rules, and then suggest the most suitable ranking method, similarity measure and uncertainty measure that should be used in the CWW paradigm. The results will be useful in understanding the uncertainties associated with linguistic terms and hence how to use them effectively in survey design and linguistic information processing.

I. Introduction

Zadeh coined the phrase "*computing with words*" (CWW) [32], [33]. According to [33], CWW is *"a methodology in which the objects of computation are words and propositions drawn from a natural language"*. Nikravesh [22] further pointed out that CWW *"is fundamentally different from the traditional expert systems which are simply tools to 'realize' an intelligent system, but are not able to process natural language which is imprecise, uncertain and partially true."*

There are at least two types of uncertainties associated with a word [23]: intra-personal uncertainty and inter-personal uncertainty. The former is explicitly pointed out by Wallsten and Budescu [23] as *"except in very special cases, all representations are vague to some degree in the minds of the originators and in the minds of the receivers,"* and they suggest to model it by type-1 fuzzy sets (T1 FSs). The latter is pointed out by Mendel [16] as *"words mean different things to different people"* and Wallsten and Budescu [23] 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 to use IT2 FSs in CWW [10], [12], [16].

A specific architecture (Fig. 1) is proposed in [11] for making judgements by CWW. It will be called a *perceptual computer*–Per-C for short. Perceptions (i.e., granulated terms, words) activate the Per-C and are also output by the Per-C; so, it is possible for a human to interact with the Per-C just using a vocabulary of words. In Fig. 1, the *encoder* [1] transforms linguistic perceptions into IT2 FSs that activate a *CWW engine*. The *decoder* [2] maps the output of the CWW engine into a word. Usually a vocabulary (codebook) is available, in which every word is modeled as an IT2 FS. The output of the CWW engine is mapped into a word (in that vocabulary) most similar to it.

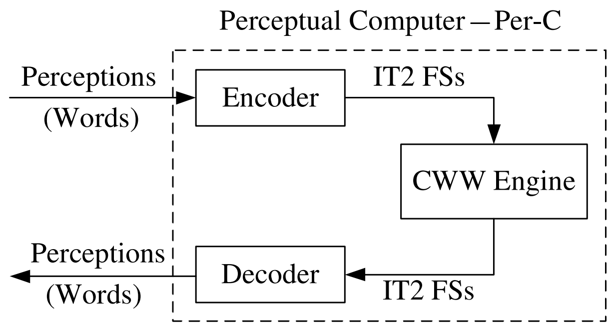

Fig. 1. Conceptual structure of CWW.

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

1) 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* [12] and the *interval end-points approach* [14], [19]. Recently, Liu and Mendel [9] proposed a new method called the *interval approach*, which captures the strong points of both the person-MF and interval end-points approaches.
2) How to construct the CWW engine, which maps IT2 FSs into IT2 FSs. There may be different kinds of CWW engines, e.g., rules [16], the linguistic weighted average[3] (LWA) [26], [27], the novel weighted average (NWA) [13], perceptual reasoning (PR) [17], [18], etc. If the CWW engine is rule-based, its output may be a crisp number (e.g., after defuzzification). On the other hand, if the CWW engine uses LWAs, NWAs and/or PR, its output will be an IT2 FS $\tilde{A}$.
3) How to map the output of the CWW engine to a word (linguistic label), i.e., the decoding problem. Mendel [11] has proposed an approach to map a crisp number into a word in the vocabulary. To map an IT2 FS to a word, it must be possible to compare the similarity between two IT2 FSs. There are four existing similarity measures for IT2 FSs in the literature [2], [4], [20], [34]. Recently, Wu and Mendel proposed a new vector similarity measure (VSM) [31].
4) How to rank IT2 FSs. In one example used in [13], we have to choose the best tactical missile system (TMS) from three competing companies. The process for evaluating each TMS is shown in Fig. 2. The overall performance of a TMS, denoted as $\tilde{Y}_i$ $(i = 1, 2, 3)$, is represented by an IT2 FS computed from several NWAs. Hence, we must be able to rank these $\tilde{Y}_i$ to find the best TMS. Only one method for ranking IT2 FSs has been proposed so far by Mitchell [21].
5) How to quantify the uncertainty associated with an IT2 FS. As pointed out by Klir [8], "*once uncertainty (and information) measures become well justified, they can very effectively be utilized for managing uncertainty*

[1] Zadeh calls this *constraint explicitation* in [32], [33]. In some of his recent talks, he calls this *precisiation*.

[2] Zadeh calls this *linguistic approximation* in [32], [33].

[3] An LWA is computed as

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

where $\tilde{X}_i$ and $\tilde{W}_i$ are words modeled by IT2 FSs.

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 [5], [8], e.g., the principle of minimum uncertainty, the principle of minimum uncertainty, and the principle of uncertainty invariance. Five uncertainty measures have been proposed in [29]; however, an open problem is which one to use.

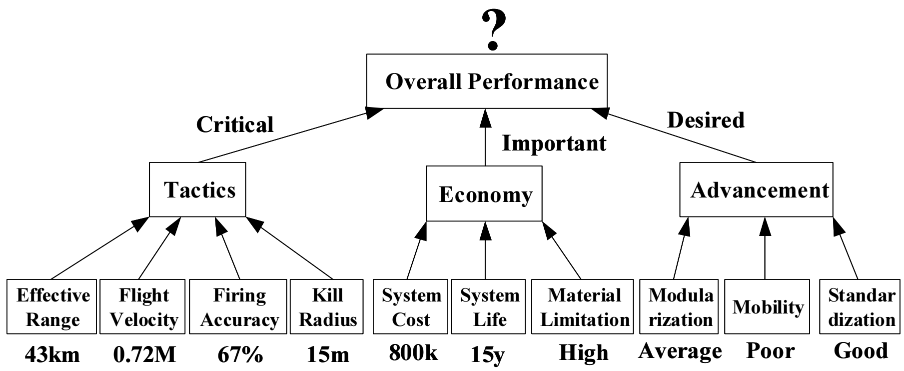

Fig. 2. Tactical missile systems evaluation process.

Problems (3)-(5) will be considered in this study. The objectives are to:

1) Evaluate ranking methods, similarity measures and uncertainty measures for IT2 FSs based on real survey data;
2) Test ranking methods and the LWA against psychological rules; and,
3) Suggest the most suitable ranking method, similarity measure and uncertainty measure that should be used in the CWW paradigm.

Our research results will also help people better understand the uncertainties associated with linguistic terms and hence how to use them effectively in survey design and linguistic information processing.

The rest of this report is organized as follows: Section II presents the 32 word FOUs used in this study. Section III proposes a new ranking method for IT2 FSs and compared it with Mitchell's method. Section IV proposes a new similarity measure for IT2 FSs and compared it with the existing five methods. Section V computes the uncertainty measures for the 32 words and studies their relationship. Section VI tests ranking methods and the LWA against psychological rules. Section VII draws conclusions.

II. Word FOUs

The dataset used herein was collected from 28 subjects at the Jet Propulsion Laboratory (JPL). 32 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*). The real survey data for the 32 words are available online [1].

Liu and Mendel's interval approach for word modeling [9] was used to map these data intervals into FOUs. After some pre-processing, during which some intervals (e.g., outliers) are eliminated, the collection of remaining intervals is classified as either an interior, left-shoulder or right-shoulder IT2 FS. Then, each of the data intervals is individually mapped into its respective T1 interior, left-shoulder or right-shoulder MF, after which the union of all of these T1 MFs is taken, and the union is upper and lower bounded. The result is an FOU for an IT2 FS, 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 obtained in Fig. 3.

III. Ranking Methods for IT2 FSs

Though there are at least 29 different methods for ranking type-1 fuzzy numbers [24], [25] in the literature, to the best knowledge of the author, only one method on ranking IT2 FSs has been published by Mitchell [21]. We will introduce Mitchell's method, point out its limitations, and also propose a new ranking method for IT2 FSs.

A. Mitchell's Method for Ranking IT2 FSs

Mitchell [21] proposed a ranking method for general type-2 FSs. When specified to IT2 FSs, the procedure is:

1) Discretize the primary variable's universe of discourse, X, into L points, that are used by all $\tilde{A}_i$.
2) Find H embedded T1 FSs, A_e^{mh}, $m = 1, \ldots, M$, $h = 1, \ldots, H$, for each of the M IT2 FSs $\tilde{A}_m$ randomly:

$$\mu_{A_e^{mh}}(x_l) = r_{mh}(x_l) \times [\overline{\mu}_{\tilde{A}_m}(x_l) - \underline{\mu}_{\tilde{A}_m}(x_l)] + \underline{\mu}_{\tilde{A}_m}(x_l) \qquad l = 1, 2, \ldots, L \tag{2}$$

   where $r_{mh}(x_l)$ is a random number chosen uniformly in $[0, 1]$, and $\underline{\mu}_{\tilde{A}_m}(x_l)$ and $\overline{\mu}_{\tilde{A}_m}(x_l)$ are the lower and upper memberships of $\tilde{A}_m$ at x_l.
3) Form the H^M different combinations of $(A_e^{1h}, A_e^{2h}, \ldots, A_e^{Mh})_i$, $i = 1, \ldots, H^M$.
4) Use a T1 FS ranking method (the centroids are used in this study) 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}.
5) Compute the final rank of $\tilde{A}_m$ as

$$r_m = \frac{1}{H^M} \sum_{i=1}^{H^M} r_{mi}, \quad m = 1, \ldots, M \tag{3}$$

Observe from the above procedure that:

1) The output ranking, r_m, is a crisp number; however, usually it is not an integer. We need to sort these r_m $(m = 1, \ldots, M)$ again to find the correct ranking.
2) 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 H is chosen as a small number 2, we have to evaluate $2^{32} = 4.295 \times 10^9$ T1 FS rankings, each of which involves 32 T1 FSs. This is highly impractical. Though two fast algorithms are proposed in [21], because our FOUs have lots of overlap, the computational cost cannot be reduced significantly.
3) Because there are random numbers involved, r_m may change from experiment to experiment. When H is large, some kind of stochastic convergence can be expected to occur (e.g., convergence in probability); however, as mentioned above, the computational cost is prohibitive.

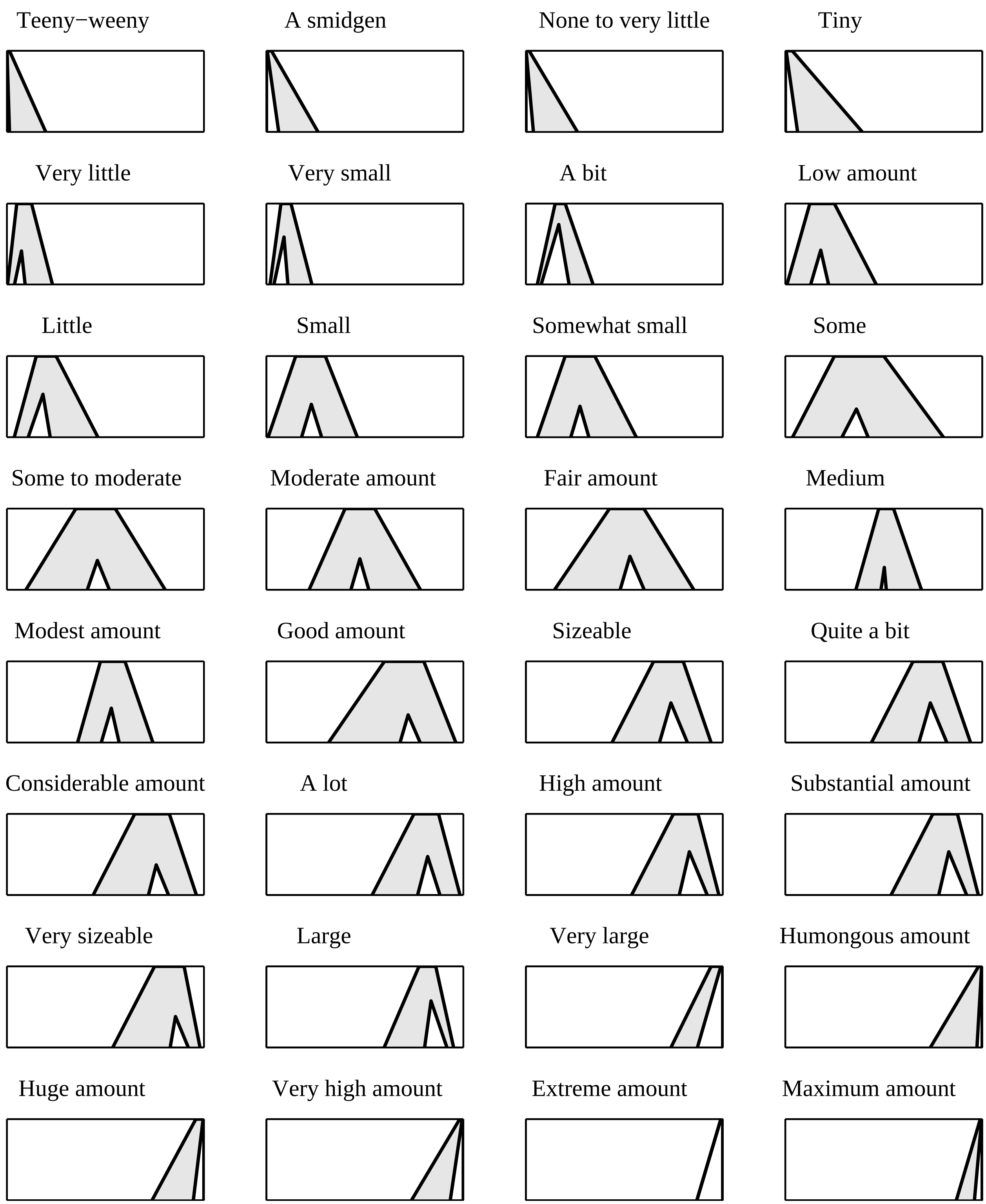

Fig. 3. The 32 word FOUs ranked by their centers of centroid. To read this figure, scan from left to right starting at the top of the page.

B. A New Centroid-Based Ranking Method

Because the high computational cost for Mitchell's IT2 FS ranking method makes it impractical to use, a simple ranking method based on the centroids of IT2 FSs is proposed in this subsection.

Definition 1: [16] 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.,

$$C(\tilde{A}) \equiv \bigcup_{\forall A_e} c(A_e) = [c_l(\tilde{A}), c_r(\tilde{A})], \tag{4}$$

where $\bigcup$ is the union operation, and

$$c_l(\tilde{A}) = \min_{\forall A_e} c(A_e) \tag{5}$$

$$c_r(\tilde{A}) = \max_{\forall A_e} c(A_e) \tag{6}$$

$$c(A_e) = \frac{\sum_{i=1}^{L} x_i \mu_{A_e}(x_i)}{\sum_{i=1}^{L} \mu_{A_e}(x_i)}. \quad \square \tag{7}$$

It has been shown [7], [15], [16] that $c_l(\tilde{A})$ and $c_r(\tilde{A})$ can be expressed as

$$c_l(\tilde{A}) = \frac{\sum_{i=1}^{L} x_i \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^{N} x_i \underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{L} \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^{N} \underline{\mu}_{\tilde{A}}(x_i)} \tag{8}$$

$$c_r(\tilde{A}) = \frac{\sum_{i=1}^{R} x_i \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^{N} x_i \overline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{R} \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^{N} \overline{\mu}_{\tilde{A}}(x_i)}. \tag{9}$$

Switch points L and R, as well as $c_l(\tilde{A})$ and $c_r(\tilde{A})$, are computed by using the iterative KM Algorithms [7], [16], [28].

In the new centroid-based ranking method, we first compute the average centroid for each IT2 FS $\tilde{A}_i$,

$$c(\tilde{A}_i) = \frac{c_l(\tilde{A}_i) + c_r(\tilde{A}_i)}{2} \tag{10}$$

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

C. Comparative Studies

The ranking of the 32 word FOUs using the centroid-based method has been presented in Fig. 3. Observe that

1) The four smallest terms are left shoulders, the six largest terms are right shoulders, and the terms in-between have interior FOUs. This order is intuitive.
2) Visual examination shows that the ranking is reasonable, and 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, we partitioned the 32 words into four groups, each with eight words, and then ranked each group individually. $H = 3$ was used. The results are shown in Fig. 4. Words which have different rank than that in Fig. 3 are shaded in red. Observe that

1) 14 words are in different order than that in Fig. 3. More words in different orders are expected if we rank the 32 words in one pass using Mitchell's method.
2) The natural transition from left shoulders to interior FOUs and then to right shoulders is destroyed. "Very large" is smaller than "large," which is counter-intuitive.

In summary, the centroid-based ranking method for IT2 FSs seems to be a better choice than Mitchell's method in CWW. We shall have more to say about the centroid-based ranking method in Section IV.

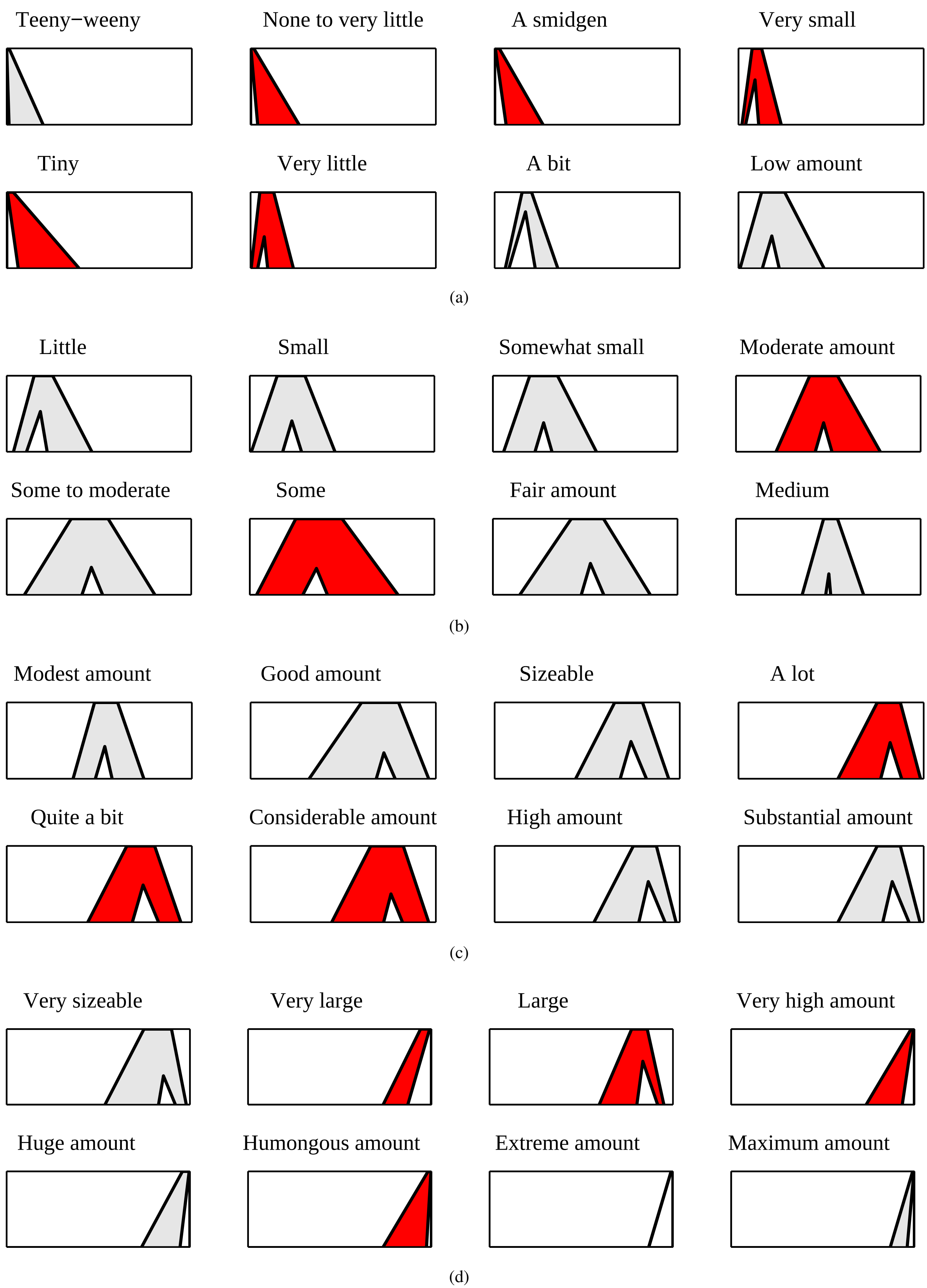

Fig. 4. Ranking of the 32 word FOUs using Mitchell's method.

IV. Similarity Measures

In this section, we briefly introduce five existing similarity measures [2], [4], [20], [31], [34] for IT2 FSs and then propose a new similarity measure with reduced computational cost. Their performance will also be compared. The following four properties [31] will serve as criteria in the comparison:

1) *Reflexivity:* $s(\tilde{A},\tilde{B}) = 1 \Leftrightarrow \tilde{A} = \tilde{B}$.
2) *Symmetry:* $s(\tilde{A},\tilde{B}) = s(\tilde{B},\tilde{A})$.
3) *Transitivity:* (a) If $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are of the same shape and $c(\tilde{A}) < c(\tilde{B}) < c(\tilde{C})$ [see the definition of $c(\tilde{A})$ in (10)], then $s(\tilde{A},\tilde{B}) > s(\tilde{A},\tilde{C}) \geq 0$, or $s(\tilde{A},\tilde{B}) = s(\tilde{A},\tilde{C}) = 0$. (b) If $c(\tilde{A}) = c(\tilde{B}) = c(\tilde{C})$ and $\tilde{A} < \tilde{B} < \tilde{C}$, then $s(\tilde{A},\tilde{B}) > s(\tilde{A},\tilde{C})$.
4) *Overlap:* If $\tilde{A} \cap \tilde{B} \neq \emptyset$, then $s(\tilde{A},\tilde{B}) > 0$.

A. Mitchell's IT2 FS Similarity Measure

Mitchell was the first to define a similarity measure for *general* T2 FSs [20]. For the purpose of this paper, only its special case is explained, when both $\tilde{A}$ and $\tilde{B}$ are IT2 FSs:

(1) Discretize the primary variable's universe of discourse, X, into L points, that are used by both $\tilde{A}$ and $\tilde{B}$.

(2) Find M embedded T1 FSs for IT2 FS $\tilde{A}$ ($m = 1, 2, \ldots, M$), i.e.

$$\mu_{A_e^m}(x_l) = r_m(x_l) \times [\overline{\mu}_{\tilde{A}}(x_l) - \underline{\mu}_{\tilde{A}}(x_l)] + \underline{\mu}_{\tilde{A}}(x_l) \qquad l = 1, 2, \ldots, L \tag{11}$$

where $r_m(x_l)$ is a random number chosen uniformly in $[0, 1]$, and $\underline{\mu}_{\tilde{A}}(x_l)$ and $\overline{\mu}_{\tilde{A}}(x_l)$ are the lower and upper memberships of $\tilde{A}$ at x_l.

(3) Similarly, find N embedded T1 MFs, $\mu_{B_e^n}$ ($n = 1, 2, \ldots, N$), for IT2 FS $\tilde{B}$, i.e.,

$$\mu_{B_e^n}(x_l) = r_n(x_l) \times [\overline{\mu}_{\tilde{B}}(x_l) - \underline{\mu}_{\tilde{B}}(x_l)] + \underline{\mu}_{\tilde{B}}(x_l) \qquad l = 1, 2, \ldots, L \tag{12}$$

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

$$s_{_M}(\tilde{A},\tilde{B}) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} s_{mn}, \tag{13}$$

where

$$s_{mn} = s(A_e^m, A_e^n) \tag{14}$$

and s_{mn} can be any T1 FS similarity measure. Jaccard's similarity measure [6] is used in this study.

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

1) 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.
2) It does not satisfy symmetry because of the random numbers.
3) $s_{_M}(\tilde{A},\tilde{B})$ may change from experiment to experiment. When both M and N are large, some kind of stochastic convergence can be expected to occur (e.g., convergence in probability); however, the computational cost is heavy because the computation of (13) requires direct enumeration of all MN embedded T1 FSs.

B. Gorzalczany's IT2 FS Compatibility Measure

Gorzalczany [4] defined the *degree of compatibility*, $s_G(\tilde{A},\tilde{B})$, between two IT2 FSs $\tilde{A}$ and $\tilde{B}$ as

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

Gorzalczany's compatibility measure also does not satisfy reflexivity. It can be shown [31] that as long as $\max_{x\in X}\underline{\mu}_{\tilde{A}}(x) = \max_{x\in X}\underline{\mu}_{\tilde{B}}(x)$ and $\max_{x\in X}\overline{\mu}_{\tilde{A}}(x) = \max_{x\in X}\overline{\mu}_{\tilde{B}}(x)$, no matter how different the shapes of $\tilde{A}$ and $\tilde{B}$ are, Gorzalczany's compatibility measure always gives $s_G(\tilde{A},\tilde{B}) = s_G(\tilde{B},\tilde{A}) = [1,1]$, which is counter-intuitive.

C. Bustince's IT2 FS Similarity Measure

Bustince's *interval-valued normal similarity measure* [2] is defined as

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

where

$$s_L(\tilde{A},\tilde{B}) = \Upsilon_L(\tilde{A},\tilde{B}) \star \Upsilon_L(\tilde{B},\tilde{A}) \tag{17}$$

and

$$s_U(\tilde{A},\tilde{B}) = \Upsilon_U(\tilde{A},\tilde{B}) \star \Upsilon_U(\tilde{B},\tilde{A}), \tag{18}$$

$\star$ can be any t-norm (e.g., minimum), and $[\Upsilon_L(\tilde{A},\tilde{B}), \Upsilon_U(\tilde{A},\tilde{B})]$ is an *interval valued inclusion grade indicator* [2] of $\tilde{A}$ in $\tilde{B}$. $\Upsilon_L(\tilde{A},\tilde{B})$ and $\Upsilon_U(\tilde{A},\tilde{B})$ used in this study (and taken from [2]) are computed as

$$\Upsilon_L(\tilde{A},\tilde{B}) = \inf_{x\in X}\left\{1, \min(1-\underline{\mu}_{\tilde{A}}(x)+\underline{\mu}_{\tilde{B}}(x), 1-\overline{\mu}_{\tilde{A}}(x)+\overline{\mu}_{\tilde{B}}(x))\right\} \tag{19}$$
$$\Upsilon_U(\tilde{A},\tilde{B}) = \inf_{x\in X}\left\{1, \max(1-\underline{\mu}_{\tilde{A}}(x)+\underline{\mu}_{\tilde{B}}(x), 1-\overline{\mu}_{\tilde{A}}(x)+\overline{\mu}_{\tilde{B}}(x))\right\} \tag{20}$$

Bustince's similarity measure does not satisfy transitivity (a), i.e., when $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are of the same shape and $c(\tilde{A}) < c(\tilde{B}) < c(\tilde{C})$, $s_B(\tilde{A},\tilde{B}) = s_B(\tilde{A},\tilde{C}) > 0$ [31].

D. Zeng and Li's IT2 FS Similarity Measure

Zeng and Li [34] proposed the following similarity measure for IT2 FSs if the universes of discourse of $\tilde{A}$ and $\tilde{B}$ are discrete:

$$s_z(\tilde{A},\tilde{B}) = 1 - \frac{1}{2n}\sum_{i=1}^{n}\left(|\underline{\mu}_{\tilde{A}}(x_i)-\underline{\mu}_{\tilde{B}}(x_i)| + |\overline{\mu}_{\tilde{A}}(x_i)-\overline{\mu}_{\tilde{B}}(x_i)|\right), \tag{21}$$

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

$$s_z(\tilde{A},\tilde{B}) = 1 - \frac{1}{2(b-a)}\int_a^b\left(|\underline{\mu}_{\tilde{A}}(x)-\underline{\mu}_{\tilde{B}}(x)| + |\overline{\mu}_{\tilde{A}}(x)-\overline{\mu}_{\tilde{B}}(x)|\right)dx. \tag{22}$$

Zeng and Li's similarity measure also does not satisfy transitivity (a) [31]. When $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are of the same shape and $c(\tilde{A}) < c(\tilde{B}) < c(\tilde{C})$, $s_z(\tilde{A},\tilde{B}) = s_z(\tilde{A},\tilde{C}) > 0$ or $s_z(\tilde{A},\tilde{B}) < s_z(\tilde{A},\tilde{C})$.

E. The Vector Similarity Measure

Recently Wu and Mendel [31] proposed a vector similarity measure (VSM), which has two components:

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

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

To compute $s_1(\tilde{A},\tilde{B})$, we first compute their centers of centroid, $c(\tilde{A})$ and $c(\tilde{B})$, and then move $\tilde{B}$ to $\tilde{B}'$ so that $c(\tilde{A}) = c(\tilde{B}')$. $s_1(\tilde{A},\tilde{B})$ is then computed as the ratio of the *average cardinalities* [see (34)] of $\tilde{A}\cap\tilde{B}'$ and $\tilde{A}\cup\tilde{B}'$, i.e.

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

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

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

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

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

A scalar similarity measure can be computed from the VSM as

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

$s_s(\tilde{A},\tilde{B})$ satisfies all four properties [31]; however, its computational cost is high, e.g., centroids of $\tilde{A}$ and $\tilde{B}$ need to be computed before $s_1(\tilde{A},\tilde{B})$ and $s_2(\tilde{A},\tilde{B})$ can be obtained.

F. Jaccard's Similarity Measure for IT2 FSs

A new similarity measure, which is an extension of Jaccard's similarity measure for T1 FSs, is proposed in this subsection. It is motivated by (24): if we compute $p(\tilde{A}\cap\tilde{B})/p(\tilde{A}\cup\tilde{B})$ directly instead of $p(\tilde{A}\cap\tilde{B}')/p(\tilde{A}\cup\tilde{B}')$, then we utilize both shape and proximity information simultaneously without computing the centroids and aligning $\tilde{A}$ and $\tilde{B}$. The new similarity measure is hence defined as

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

G. Comparative Studies

The similarities among the 32 word FOUs computed using the above six measures are shown in Tables I-VI, respectively. Observe that:

1) Generally Mitchell's method gives $s_M(\tilde{A}, \tilde{A}) < 1$ (see the diagonal terms in Table I), except for the special case that $\tilde{A}$ is a T1 FS (see the (31,31)th term in Table I). Also, because $s_M(\tilde{A}, \tilde{B}) \neq s_M(\tilde{B}, \tilde{A})$, the matrix is not symmetric.
2) Gorzalczany's method indicates "very large," "humongous amount," "huge amount," "very high amount," "extreme amount" and "maximum amount" are equivalent (see the block of ones at the bottom-right corner of Table II), which is counter-intuitive.
3) The second column of Table III shows that as a word becomes larger, its similarity to "teeny-weeny" does not decrease monotonically as computed by Bustince's similarity measure, which is counter-intuitive.
4) Zeng and Li's method gives large similarity whether $\tilde{A}$ and $\tilde{B}$ overlap or not (see Table IV), and the similarity may increase as two words get further away from each other, which is again counter-intuitive.
5) The VSM gives much more reasonable results (see Table V). Generally the similarity decreases monotonically as two words gets further away. This phenomena also indicates that the centroid-based ranking method is reasonable. However, observe that some diagonal terms are 0.99 instead of 1 because of the errors in computing the centroid and aligning the FSs.
6) Jaccard's similarity measure gives similar results as the VSM, but they are more accurate (e.g., the diagonal terms in Table VI are all 1). Also, simulations show that Jaccard's method is about 3.5 times faster than the VSM.
7) Except for Mitchell's method, all other similarity measures indicate that "sizable" and "quite a bit" are equivalent, and "high amount" and "substantial amount" are equivalent.

In summary, Jaccard's similarity measure seems to be the best for CWW.

It is also interesting to connect similarity measures to ranking methods. Because the terms in Table VI are ranked in ascending order according to the centroid-based ranking method, we would expect a monotonic increase in the similarity as the rankings of two terms get closer, e.g., the similarity between the ith term and the jth term should increase as $|i-j|$ decreases. Observe from Table VI that the above conjecture is true for most cases; however, there are some counterexamples, e.g., the similarity between "a smidgen" and "teeny-weeny" is smaller than that between "none to very little" and "teeny-weeny." This may suggest that we should swap the positions of "a smidgen" and "none to very little;" however, the swap cannot solve the problem completely, e.g., after the swap, the similarities between them and "very little" become a counterexample. For the 32 terms used in this report, it is impossible to make the ranking completely coincide with the similarities; so, for simplicity, we stick to the centroid-based ranking method and do not fine-tune it using a similarity measure.

V. Uncertainty Measures

Wu and Mendel [29] proposed five uncertainty measures for IT2 FSs: *centroid, cardinality, fuzziness, variance* and *skewness*; however, an open problem is which one to use. In this section, we will try to answer this question by distinguishing between two types of uncertainties [23], *intra-personal uncertainty* and *inter-personal uncertainty*, and studying which uncertainty measure best captures both of them.

TABLE I

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. Teeny-weeny	.54	.50	.50	.39	.35	.31	.12	.14	.11	.12	.06	.04	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. A smidgen	.50	.65	.60	.53	.40	.35	.19	.19	.16	.17	.10	.08	.04	0	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. None to very little	.50	.62	.60	.51	.39	.37	.19	.20	.17	.17	.10	.07	.04	0	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.37	.54	.50	.61	.37	.33	.30	.31	.29	.29	.19	.14	.09	.04	.06	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very little	.35	.41	.40	.37	.55	.54	.26	.26	.21	.20	.11	.08	.03	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very small	.31	.35	.36	.32	.54	.61	.27	.26	.22	.21	.11	.08	.03	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.11	.19	.20	.31	.26	.27	.70	.46	.51	.41	.30	.18	.10	.03	.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8. Low amount	.13	.20	.20	.31	.25	.26	.47	.55	.54	.53	.39	.26	.18	.10	.12	.03	.02	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9. Little	.10	.16	.16	.29	.20	.22	.50	.54	.58	.53	.42	.28	.19	.11	.13	.03	.03	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10. Small	.12	.17	.17	.28	.20	.21	.42	.54	.53	.55	.45	.30	.20	.12	.14	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11. Somewhat small	.06	.10	.10	.19	.11	.11	.30	.40	.43	.45	.54	.39	.31	.21	.22	.10	.09	.07	.02	.02	.02	0	0	0	0	0	0	0	0	0	0	0
12. Some	.04	.07	.08	.14	.08	.08	.18	.27	.28	.30	.40	.53	.49	.41	.42	.24	.28	.21	.14	.14	.14	.07	.07	.07	.07	.05	0	0	.01	0	0	0
13. Some to moderate	.02	.04	.04	.09	.03	.04	.10	.18	.18	.20	.31	.49	.54	.47	.49	.30	.36	.26	.17	.17	.16	.09	.08	.08	.09	.06	.01	.01	.01	.01	0	0
14. Moderate amount	0	0	0	.04	0	0	.03	.10	.11	.12	.21	.40	.48	.55	.48	.38	.41	.27	.18	.17	.17	.09	.08	.09	.09	.06	0	0	0	0	0	0
15. Fair amount	0	.02	.02	.06	.01	.01	.06	.12	.12	.14	.22	.43	.49	.49	.54	.33	.39	.34	.24	.24	.24	.13	.13	.13	.13	.09	.02	.02	.02	.02	0	0
16. Medium	0	0	0	0	0	0	0	.03	.03	.03	.11	.26	.31	.37	.34	.51	.47	.24	.15	.15	.15	.05	.06	.05	.05	.03	0	0	0	0	0	0
17. Modest amount	0	0	0	0	0	0	0	.02	.03	.03	.10	.28	.36	.43	.40	.48	.56	.30	.21	.21	.20	.09	.09	.09	.09	.05	0	0	0	0	0	0
18. Good amount	0	0	0	.01	0	0	0	.02	.02	.03	.07	.21	.25	.27	.33	.24	.30	.53	.50	.49	.48	.35	.36	.37	.34	.31	.10	.09	.10	.09	.03	.02
19. Sizeable	0	0	0	0	0	0	0	0	0	0	.02	.14	.17	.18	.23	.15	.21	.50	.57	.59	.55	.43	.41	.42	.40	.37	.10	.10	.10	.09	.02	.02
20. Quite a bit	0	0	0	0	0	0	0	0	0	0	.02	.14	.17	.18	.25	.15	.21	.49	.58	.57	.54	.42	.44	.42	.40	.38	.10	.09	.09	.09	.03	.02
21. Considerable amount	0	0	0	0	0	0	0	0	0	0	.02	.13	.17	.18	.23	.15	.20	.47	.53	.53	.53	.45	.45	.45	.43	.40	.13	.12	.12	.11	.04	.03
22. A lot	0	0	0	0	0	0	0	0	0	0	0	.07	.09	.09	.13	.06	.09	.36	.43	.43	.44	.57	.57	.58	.53	.52	.20	.19	.20	.18	.09	.06
23. High amount	0	0	0	0	0	0	0	0	0	0	0	.07	.09	.08	.12	.05	.09	.36	.42	.42	.45	.58	.59	.59	.56	.55	.21	.19	.21	.18	.08	.06
24. Substantial amount	0	0	0	0	0	0	0	0	0	0	0	.07	.09	.09	.13	.06	.09	.37	.43	.41	.45	.58	.59	.60	.55	.54	.21	.19	.20	.19	.09	.06
25. Very sizeable	0	0	0	0	0	0	0	0	0	0	0	.07	.08	.09	.12	.06	.09	.35	.40	.40	.41	.54	.56	.57	.54	.51	.23	.20	.21	.20	.10	.07
26. Large	0	0	0	0	0	0	0	0	0	0	0	.05	.06	.06	.10	.03	.05	.31	.37	.37	.39	.53	.54	.54	.51	.60	.20	.20	.19	.18	.06	.04
27. Very large	0	0	0	0	0	0	0	0	0	0	0	0	.01	0	.02	0	0	.10	.10	.10	.12	.20	.21	.21	.24	.19	.76	.57	.64	.66	.61	.39
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	0	.02	0	0	.09	.10	.09	.12	.19	.20	.19	.21	.19	.57	.57	.58	.58	.51	.44
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	0	.02	0	0	.10	.10	.10	.13	.20	.20	.20	.22	.18	.64	.58	.62	.64	.57	.46
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	0	.01	0	.02	0	0	.09	.09	.10	.11	.18	.19	.19	.21	.18	.64	.58	.64	.67	.61	.47
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.03	.03	.02	.04	.09	.09	.09	.10	.06	.61	.51	.56	.61	1	.65
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	.05	.40	.43	.45	.47	.65	.69

TABLE II

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. Teeny-weeny	1	.98	.98	.96	.42	.37	.25	.31	.27	.29	.21	.20	.11	0	.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. A smidgen	.98	1	1	1	.53	.49	.31	.36	.32	.33	.27	.25	.17	.05	.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. None to very little	.98	1	1	1	.44	.41	.30	.35	.32	.33	.26	.24	.17	.05	.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.96	1	1	1	.56	.53	.37	.41	.38	.39	.33	.31	.24	.16	.19	.03	.03	.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very little	.42	.64	.44	.67	.99	.99	.52	.50	.44	.45	.35	.30	.19	.02	.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very small	.37	.55	.41	.58	.85	.99	.60	.50	.45	.45	.35	.30	.19	.02	.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.25	.31	.30	.37	.48	.57	1	.78	.85	.64	.50	.42	.31	.19	.23	0	0	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8. Low amount	.31	.36	.35	.41	.50	.50	.99	.99	.99	.72	.50	.50	.39	.31	.32	.15	.15	.15	.02	.02	.02	0	0	0	0	0	0	0	0	0	0	0
9. Little	.27	.32	.32	.38	.44	.45	.99	.89	.99	.70	.50	.50	.39	.31	.32	.15	.15	.14	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0
10. Small	.29	.33	.33	.39	.45	.45	.76	.73	.76	1	.75	.50	.44	.35	.36	.18	.18	.16	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0
11. Somewhat small	.21	.27	.26	.33	.35	.35	.50	.50	.50	.76	1	.62	.50	.44	.42	.31	.31	.25	.15	.15	.14	.03	.02	.02	.02	0	0	0	0	0	0	0
12. Some	.20	.25	.24	.31	.30	.30	.42	.50	.50	.50	.63	.99	.55	.50	.50	.50	.50	.41	.35	.35	.35	.26	.26	.26	.26	.21	.06	.06	.06	.06	0	0
13. Some to moderate	.11	.17	.17	.24	.19	.19	.31	.39	.39	.44	.50	.54	.99	.93	.69	.70	.69	.45	.39	.39	.39	.28	.28	.28	.28	.24	.07	.07	.07	.07	0	0
14. Moderate amount	0	.05	.05	.16	.02	.02	.19	.31	.31	.35	.44	.50	.91	1	.73	.73	.73	.45	.38	.38	.38	.27	.27	.27	.27	.22	.05	.05	.05	.05	0	0
15. Fair amount	.05	.11	.11	.19	.11	.11	.23	.32	.32	.36	.42	.50	.66	.71	.99	.74	.99	.50	.45	.45	.45	.34	.34	.34	.34	.29	.12	.12	.12	.12	0	0
16. Medium	0	0	0	.03	0	0	0	.15	.15	.18	.31	.50	.76	.82	.87	1	.88	.44	.36	.36	.36	.20	.22	.22	.22	.14	0	0	0	0	0	0
17. Modest amount	0	0	0	.03	0	0	0	.15	.15	.18	.31	.50	.66	.70	.98	.75	.99	.50	.43	.43	.43	.28	.29	.29	.29	.22	0	0	0	0	0	0
18. Good amount	0	0	0	.06	0	0	.03	.15	.14	.16	.25	.41	.45	.45	.50	.44	.50	1	1	1	.81	.56	.50	.50	.50	.50	.30	.27	.29	.27	.16	.16
19. Sizeable	0	0	0	0	0	0	0	.02	.03	.03	.15	.35	.39	.38	.45	.36	.43	.85	1	1	.88	.69	.65	.65	.50	.58	.29	.26	.28	.26	.14	.14
20. Quite a bit	0	0	0	0	0	0	0	.02	.03	.03	.15	.35	.39	.38	.45	.36	.43	.85	1	1	.88	.69	.65	.65	.50	.58	.29	.26	.28	.26	.14	.14
21. Considerable amount	0	0	0	0	0	0	0	.02	.03	.03	.14	.35	.39	.38	.45	.36	.43	.79	1	1	1	.75	.70	.70	.50	.60	.33	.29	.31	.29	.18	.18
22. A lot	0	0	0	0	0	0	0	0	0	0	.03	.26	.28	.27	.34	.22	.28	.54	.69	.69	.69	.99	.97	.97	.77	.93	.42	.35	.37	.35	.29	.25
23. High amount	0	0	0	0	0	0	0	0	0	0	.02	.26	.28	.27	.34	.22	.29	.50	.64	.64	.64	.92	1	1	.85	.99	.55	.35	.37	.35	.41	.25
24. Substantial amount	0	0	0	0	0	0	0	0	0	0	.02	.26	.28	.27	.34	.22	.29	.50	.64	.64	.64	.92	1	1	.85	.99	.55	.35	.37	.35	.41	.25
25. Very sizeable	0	0	0	0	0	0	0	0	0	0	.02	.26	.28	.27	.34	.22	.29	.50	.50	.50	.50	.84	.99	.99	.99	.99	.64	.37	.40	.37	.50	.27
26. Large	0	0	0	0	0	0	0	0	0	0	0	.21	.24	.22	.29	.14	.22	.50	.57	.57	.56	.86	.95	.95	.83	.99	.51	.32	.34	.32	.35	.19
27. Very large	0	0	0	0	0	0	0	0	0	0	0	.06	.07	.05	.12	0	0	.30	.29	.29	.33	.41	.47	.47	.51	.44	1	1	1	1	1	1
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	.06	.07	.05	.12	0	0	.27	.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	.06	.07	.05	.12	0	0	.29	.28	.28	.31	.37	.37	.37	.40	.34	1	1	1	1	1	1
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	.06	.07	.05	.12	0	0	.27	.26	.26	.29	.35	.35	.35	.37	.32	1	1	1	1	1	1
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.16	.14	.14	.18	.27	.33	.33	.36	.28	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	.16	.14	.14	.18	.25	.25	.25	.27	.19	1	1	1	1	1	1

TABLE III

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. Teeny-weeny	1	.52	.62	.49	.01	.01	.03	.06	.06	.06	.08	.13	.13	.13	.14	.11	.12	.16	.16	.16	.16	.17	.17	.17	.17	.16	0	.06	.01	0	0	.03
2. A smidgen	.52	1	.79	.82	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	0	0
3. None to very little	.62	.79	1	.75	0	0	0	0	0	0	0	.02	.02	.02	.02	.01	.01	.03	.03	.03	.03	.03	.03	.03	.03	.03	0	.03	.01	0	0	.03
4. Tiny	.49	.82	.75	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	0	0
5. Very little	.01	0	0	0	1	.74	.23	.29	.24	.30	.30	.30	.30	.30	.30	.30	.29	.30	.26	.26	.30	.27	.24	.24	.31	.22	0	.06	.01	0	0	.03
6. Very small	.01	0	0	0	.74	1	.32	.29	.22	.26	.21	.21	.21	.22	.21	.21	.22	.21	.22	.22	.21	.23	.23	.23	.23	.21	0	.06	.01	0	0	.03
7. A bit	.03	0	0	0	.23	.32	1	.72	.72	.48	.31	.27	.14	.14	.14	.14	.13	.14	.13	.13	.14	.14	.14	.14	.14	.14	0	.06	.01	0	0	.03
8. Low amount	.06	0	0	0	.29	.29	.72	1	.87	.75	.60	.32	.29	.29	.30	.29	.29	.29	.26	.26	.29	.27	.24	.24	.30	.22	0	.06	.01	0	0	.03
9. Little	.06	0	0	0	.24	.22	.72	.87	1	.71	.58	.30	.24	.24	.25	.24	.24	.24	.24	.24	.24	.24	.23	.23	.24	.22	0	.06	.01	0	0	.03
10. Small	.06	0	0	0	.30	.26	.48	.75	.71	1	.73	.33	.30	.30	.30	.30	.29	.30	.26	.26	.30	.27	.24	.24	.30	.22	0	.06	.01	0	0	.03
11. Somewhat small	.08	0	0	0	.30	.21	.31	.60	.58	.73	1	.51	.35	.31	.30	.32	.29	.31	.26	.26	.31	.27	.24	.24	.31	.22	0	.06	.01	0	0	.03
12. Some	.13	0	.02	0	.30	.21	.27	.32	.30	.33	.51	1	.74	.44	.55	.33	.33	.33	.26	.26	.32	.27	.23	.23	.32	.22	0	.06	.01	0	0	.03
13. Some to moderate	.13	0	.02	0	.30	.21	.14	.29	.24	.30	.35	.74	1	.76	.74	.38	.38	.33	.26	.26	.32	.26	.24	.24	.31	.22	0	.05	.01	0	0	.02
14. Moderate amount	.13	0	.02	0	.30	.22	.14	.29	.24	.30	.31	.44	.76	1	.74	.52	.52	.31	.26	.26	.31	.27	.24	.24	.32	.22	0	.04	0	0	0	.02
15. Fair amount	.14	0	.02	0	.30	.21	.14	.30	.25	.30	.30	.55	.74	.74	1	.54	.62	.44	.26	.26	.30	.27	.24	.24	.30	.22	0	.05	.01	0	0	.02
16. Medium	.11	0	.01	0	.30	.21	.14	.29	.24	.30	.32	.33	.38	.52	.54	1	.75	.33	.26	.26	.32	.27	.24	.24	.31	.22	0	.03	0	0	0	.01
17. Modest amount	.12	0	.01	0	.29	.22	.13	.29	.24	.29	.29	.33	.38	.52	.62	.75	1	.33	.26	.26	.29	.27	.24	.24	.29	.22	0	.03	0	0	0	.01
18. Good amount	.16	0	.03	0	.30	.21	.14	.29	.24	.30	.31	.33	.33	.31	.44	.33	.33	1	.79	.79	.76	.48	.43	.43	.43	.32	0	0	0	0	0	0
19. Sizeable	.16	0	.03	0	.26	.22	.13	.26	.24	.26	.26	.26	.26	.26	.26	.26	.26	.79	1	1	.84	.59	.55	.55	.46	.42	0	.02	0	0	0	0
20. Quite a bit	.16	0	.03	0	.26	.22	.13	.26	.24	.26	.26	.26	.26	.26	.26	.26	.26	.79	1	1	.84	.59	.55	.55	.46	.42	0	.02	0	0	0	0
21. Considerable amount	.16	0	.03	0	.30	.21	.14	.29	.24	.30	.31	.32	.32	.31	.30	.32	.29	.76	.84	.84	1	.66	.59	.59	.57	.50	0	.02	0	0	0	0
22. A lot	.17	0	.03	0	.27	.23	.14	.27	.24	.27	.27	.27	.26	.27	.27	.27	.27	.48	.59	.59	.66	1	.88	.88	.69	.83	0	.01	0	0	0	0
23. High amount	.17	0	.03	0	.24	.23	.14	.24	.23	.24	.24	.23	.24	.24	.24	.24	.24	.43	.55	.55	.59	.88	1	1	.74	.85	0	.01	0	0	0	0
24. Substantial amount	.17	0	.03	0	.24	.23	.14	.24	.23	.24	.24	.23	.24	.24	.24	.24	.24	.43	.55	.55	.59	.88	1	1	.74	.85	0	.01	0	0	0	0
25. Very sizeable	.17	0	.03	0	.31	.23	.14	.30	.24	.30	.31	.32	.31	.32	.30	.31	.29	.43	.46	.46	.57	.69	.74	.74	1	.75	0	.01	0	0	0	0
26. Large	.16	0	.03	0	.22	.21	.14	.22	.22	.22	.22	.22	.22	.22	.22	.22	.22	.32	.42	.42	.50	.83	.85	.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	.64	.65	.68	.40
28. Humongous amount	.06	0	.03	0	.06	.06	.06	.06	.06	.06	.06	.06	.05	.04	.05	.03	.03	0	.02	.02	.02	.01	.01	.01	.01	0	.49	1	.69	.68	.57	.73
29. Huge amount	.01	0	.01	0	.01	.01	.01	.01	.01	.01	.01	.01	.01	0	.01	0	0	0	0	0	0	0	0	0	0	0	.64	.69	1	.89	.68	.58
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	.65	.68	.89	1	.73	.62
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	.68	.57	.68	.73	1	.62
32. Maximum amount	.03	0	.03	0	.03	.03	.03	.03	.03	.03	.03	.03	.02	.02	.02	.01	.01	0	0	0	0	0	0	0	0	0	.40	.73	.58	.62	.62	1

TABLE IV

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

TABLE V

SIMILARITY MATRIX WHEN THE VSM [31] 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. Teeny-weeny	.99	.65	.64	.32	.34	.31	.11	.09	.09	.07	.05	.04	.02	.01	.02	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. A smidgen	.65	.99	.91	.46	.46	.40	.17	.14	.13	.11	.07	.05	.03	.02	.03	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. None to very little	.64	.91	.99	.47	.46	.40	.17	.14	.13	.11	.07	.05	.03	.02	.03	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.32	.46	.47	.99	.59	.49	.30	.28	.27	.23	.14	.09	.06	.04	.05	.02	.02	.01	.01	.01	.01	0	0	0	0	0	0	0	0	0	0	0
5. Very little	.34	.46	.46	.59	1	.78	.28	.19	.19	.15	.10	.06	.04	.03	.03	0	.01	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very small	.31	.40	.40	.49	.78	1	.28	.18	.18	.14	.09	.06	.04	.02	.03	0	.01	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.11	.17	.17	.30	.28	.28	.99	.42	.44	.33	.20	.10	.08	.05	.06	.01	.01	.02	.01	.01	.01	0	0	0	0	0	0	0	0	0	0	0
8. Low amount	.09	.14	.14	.28	.19	.18	.42	1	.88	.77	.45	.21	.16	.11	.12	.05	.05	.04	.02	.02	.02	.01	.01	.01	.01	0	0	0	0	0	0	0
9. Little	.09	.13	.13	.27	.19	.18	.44	.88	1	.72	.42	.20	.15	.10	.11	.04	.04	.04	.02	.02	.02	.01	.01	.01	.01	0	0	0	0	0	0	0
10. Small	.07	.11	.11	.23	.15	.14	.33	.77	.72	1	.55	.25	.19	.14	.14	.05	.06	.05	.02	.02	.02	.01	.01	.01	.01	0	0	0	0	0	0	0
11. Somewhat small	.05	.07	.07	.14	.10	.09	.20	.45	.42	.55	1	.38	.29	.23	.22	.10	.11	.07	.04	.04	.04	.02	.02	.02	.02	.01	0	0	0	0	0	0
12. Some	.04	.05	.05	.09	.06	.06	.10	.21	.20	.25	.38	1	.66	.43	.47	.21	.23	.20	.12	.12	.12	.07	.07	.07	.07	.05	.02	.02	.02	.02	.01	.01
13. Some to moderate	.02	.03	.03	.06	.04	.04	.08	.16	.15	.19	.29	.66	1	.64	.69	.31	.34	.26	.16	.16	.17	.09	.09	.09	.09	.06	.02	.02	.02	.02	.01	.01
14. Moderate amount	.01	.02	.02	.04	.03	.02	.05	.11	.10	.14	.23	.43	.64	1	.71	.46	.49	.27	.18	.18	.18	.09	.09	.09	.09	.06	.02	.02	.02	.02	.01	.01
15. Fair amount	.02	.03	.03	.05	.03	.03	.06	.12	.11	.14	.22	.47	.69	.71	1	.44	.49	.37	.23	.23	.24	.13	.13	.13	.13	.09	.03	.03	.03	.03	.01	.01
16. Medium	0	0	0	.02	0	0	.01	.05	.04	.05	.10	.21	.31	.46	.44	1	.64	.19	.13	.13	.12	.06	.06	.06	.05	.04	.01	.01	.01	.01	0	0
17. Modest amount	0	0	0	.02	.01	.01	.01	.05	.04	.06	.11	.23	.34	.49	.49	.64	1	.27	.19	.19	.18	.09	.09	.09	.09	.06	.01	.01	.01	.01	0	0
18. Good amount	.01	.01	.01	.01	.01	.01	.02	.04	.04	.05	.07	.20	.26	.27	.37	.19	.27	1	.59	.59	.59	.32	.30	.30	.31	.22	.08	.07	.07	.06	.03	.03
19. Sizeable	0	0	0	.01	0	0	.01	.02	.02	.02	.04	.12	.16	.18	.23	.13	.19	.59	1	1	.87	.48	.44	.44	.45	.31	.09	.08	.08	.07	.03	.03
20. Quite a bit	0	0	0	.01	0	0	.01	.02	.02	.02	.04	.12	.16	.18	.23	.13	.19	.59	1	1	.87	.48	.44	.44	.45	.31	.09	.08	.08	.07	.03	.03
21. Considerable amount	0	0	0	.01	0	0	.01	.02	.02	.02	.04	.12	.17	.18	.24	.12	.18	.59	.87	.87	1	.49	.45	.45	.47	.32	.10	.08	.09	.08	.03	.03
22. A lot	0	0	0	0	0	0	0	.01	.01	.01	.02	.07	.09	.09	.13	.06	.09	.32	.48	.48	.49	1	.85	.85	.84	.63	.16	.14	.14	.13	.06	.05
23. High amount	0	0	0	0	0	0	0	.01	.01	.01	.02	.07	.09	.09	.13	.06	.09	.30	.44	.44	.45	.85	1	1	.94	.69	.18	.15	.16	.14	.06	.05
24. Substantial amount	0	0	0	0	0	0	0	.01	.01	.01	.02	.07	.09	.09	.13	.06	.09	.30	.44	.44	.45	.85	1	1	.94	.69	.18	.15	.16	.14	.06	.05
25. Very sizeable	0	0	0	0	0	0	0	.01	.01	.01	.02	.07	.09	.09	.13	.05	.09	.31	.45	.45	.47	.84	.94	.94	1	.67	.17	.15	.15	.14	.06	.05
26. Large	0	0	0	0	0	0	0	0	0	0	.01	.05	.06	.06	.09	.04	.06	.22	.31	.31	.32	.63	.69	.69	.67	1	.19	.16	.17	.15	.06	.05
27. Very large	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.02	.03	.01	.01	.08	.09	.09	.10	.16	.18	.18	.17	.19	.99	.86	.87	.77	.29	.24
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.02	.03	.01	.01	.07	.08	.08	.08	.14	.15	.15	.15	.16	.86	.99	.81	.81	.33	.27
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.02	.03	.01	.01	.07	.08	.08	.09	.14	.16	.16	.15	.17	.87	.81	.99	.89	.34	.28
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.02	.03	.01	.01	.06	.07	.07	.08	.13	.14	.14	.14	.15	.77	.81	.89	.99	.38	.30
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	.01	.01	0	0	.03	.03	.03	.03	.06	.06	.06	.06	.06	.29	.33	.34	.38	.99	.48
32. Maximum amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	.01	.01	0	0	.03	.03	.03	.03	.05	.05	.05	.05	.05	.24	.27	.28	.30	.48	.99

TABLE VI

SIMILARITY MATRIX WHEN THE NEW 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. Teeny-weeny	1	.62	.69	.44	.43	.34	.12	.17	.12	.15	.07	.05	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. A smidgen	.62	1	.91	.71	.53	.44	.19	.23	.19	.21	.12	.09	.05	0	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3. None to very little	.69	.91	1	.65	.54	.45	.20	.24	.19	.21	.12	.09	.05	0	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Tiny	.44	.71	.65	1	.51	.43	.34	.39	.33	.35	.24	.17	.11	.05	.07	0	0	.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Very little	.43	.53	.54	.51	1	.83	.25	.31	.23	.26	.13	.09	.04	0	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Very small	.34	.44	.45	.43	.83	1	.27	.32	.24	.27	.14	.09	.04	0	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. A bit	.12	.19	.20	.34	.25	.27	1	.54	.62	.46	.34	.21	.12	.04	.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8. Low amount	.17	.23	.24	.39	.31	.32	.54	1	.88	.80	.53	.34	.22	.13	.15	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9. Little	.12	.19	.19	.33	.23	.24	.62	.88	1	.77	.55	.35	.23	.13	.15	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10. Small	.15	.21	.21	.35	.26	.27	.46	.80	.77	1	.63	.38	.25	.14	.17	.04	.03	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11. Somewhat small	.07	.12	.12	.24	.13	.14	.34	.53	.55	.63	1	.55	.40	.26	.28	.13	.11	.08	.03	.03	.03	0	0	0	0	0	0	0	0	0	0	0
12. Some	.05	.09	.09	.17	.09	.09	.21	.34	.35	.38	.55	1	.74	.60	.58	.36	.38	.26	.17	.17	.17	.09	.09	.09	.09	.06	.01	.01	.01	.01	0	0
13. Some to moderate	.02	.05	.05	.11	.04	.04	.12	.22	.23	.25	.40	.74	1	.78	.72	.43	.49	.32	.21	.21	.21	.11	.11	.11	.11	.07	.01	.01	.01	.01	0	0
14. Moderate amount	0	0	0	.05	0	0	.04	.13	.13	.14	.26	.60	.78	1	.74	.55	.60	.35	.22	.22	.22	.11	.11	.11	.11	.07	0	0	0	0	0	0
15. Fair amount	0	.02	.02	.07	.02	.02	.07	.15	.15	.17	.28	.58	.72	.74	1	.44	.59	.42	.29	.29	.29	.16	.16	.16	.16	.12	.02	.02	.02	.02	0	0
16. Medium	0	0	0	0	0	0	0	.03	.03	.04	.13	.36	.43	.55	.44	1	.76	.33	.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	.11	.38	.49	.60	.59	.76	1	.41	.26	.26	.25	.11	.11	.11	.11	.07	0	0	0	0	0	0
18. Good amount	0	0	0	.01	0	0	0	.03	.03	.03	.08	.26	.32	.35	.42	.33	.41	1	.75	.75	.74	.46	.45	.45	.46	.39	.12	.12	.12	.11	.03	.03
19. Sizeable	0	0	0	0	0	0	0	0	0	0	.03	.17	.21	.22	.29	.19	.26	.75	1	1	.90	.53	.51	.51	.51	.43	.12	.12	.12	.11	.02	.02
20. Quite a bit	0	0	0	0	0	0	0	0	0	0	.03	.17	.21	.22	.29	.19	.26	.75	1	1	.90	.53	.51	.51	.51	.43	.12	.12	.12	.11	.02	.02
21. Considerable amount	0	0	0	0	0	0	0	0	0	0	.03	.17	.21	.22	.29	.19	.25	.74	.90	.90	1	.60	.58	.58	.58	.50	.14	.15	.15	.14	.03	.04
22. A lot	0	0	0	0	0	0	0	0	0	0	0	.09	.11	.11	.16	.07	.11	.46	.53	.53	.60	1	.94	.94	.87	.72	.22	.23	.23	.22	.07	.08
23. High amount	0	0	0	0	0	0	0	0	0	0	0	.09	.11	.11	.16	.07	.11	.45	.51	.51	.58	.94	1	1	.90	.77	.22	.22	.22	.21	.08	.07
24. Substantial amount	0	0	0	0	0	0	0	0	0	0	0	.09	.11	.11	.16	.07	.11	.45	.51	.51	.58	.94	1	1	.90	.77	.22	.22	.22	.21	.08	.07
25. Very sizeable	0	0	0	0	0	0	0	0	0	0	0	.09	.11	.11	.16	.07	.11	.46	.51	.51	.58	.87	.90	.90	1	.72	.25	.25	.25	.23	.09	.08
26. Large	0	0	0	0	0	0	0	0	0	0	0	.06	.07	.07	.12	.03	.07	.39	.43	.43	.50	.72	.77	.77	.72	1	.21	.21	.21	.19	.06	.05
27. Very large	0	0	0	0	0	0	0	0	0	0	0	.01	.01	0	.02	0	0	.12	.12	.12	.14	.22	.22	.22	.25	.21	1	.67	.78	.76	.59	.38
28. Humongous amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	0	.02	0	0	.12	.12	.12	.15	.23	.22	.22	.25	.21	.67	1	.85	.88	.39	.51
29. Huge amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	0	.02	0	0	.12	.12	.12	.15	.23	.22	.22	.25	.21	.78	.85	1	.90	.43	.49
30. Very high amount	0	0	0	0	0	0	0	0	0	0	0	.01	.01	0	.02	0	0	.11	.11	.11	.14	.22	.21	.21	.23	.19	.76	.88	.90	1	.49	.51
31. Extreme amount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.03	.02	.02	.03	.07	.08	.08	.09	.06	.59	.39	.43	.49	1	.64
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	.07	.07	.08	.05	.38	.51	.49	.51	.64	1

A. Cardinality of an IT2 FS

In [29] a *normalized cardinality* for a T1 FS is defined by discretizing De Luca and Termini's cardinality definition in the continuous domain [3], $\int_X \mu_A(x)dx$, i.e.

$$p(A) = \frac{|X|}{N}\sum_{i=1}^{N}\mu_A(x_i). \tag{28}$$

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

Definition 2: [29] The cardinality of an IT2 FS $\tilde{A}$ is the union of all cardinalities of its embedded T1 FSs A_e, i.e.,

$$P(\tilde{A}) \equiv \bigcup_{\forall A_e} p(A_e) = \left[p_l(\tilde{A}), p_r(\tilde{A})\right], \tag{29}$$

where

$$p_l(\tilde{A}) = \min_{\forall A_e} p(A_e) \tag{30}$$

$$p_r(\tilde{A}) = \max_{\forall A_e} p(A_e). \quad \square \tag{31}$$

Theorem 1: [29] $p_l(\tilde{A})$ and $p_r(\tilde{A})$ in (30) and (31) can be computed as

$$p_l(\tilde{A}) = p(\underline{\mu}_{\tilde{A}}(x)) \tag{32}$$

$$p_r(\tilde{A}) = p(\overline{\mu}_{\tilde{A}}(x)). \tag{33}$$

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

$$p(\tilde{A}) = \frac{p(\underline{\mu}_{\tilde{A}}(x)) + p(\overline{\mu}_{\tilde{A}}(x))}{2}. \tag{34}$$

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

B. Fuzziness (Entropy) of an IT2 FS

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

Definition 3: [29] The fuzziness $F(\tilde{A})$ of an IT2 FS $\tilde{A}$ is the union of the fuzziness of all its embedded T1 FSs A_e, i.e.,

$$F(\tilde{A}) \equiv \bigcup_{\forall A_e} f(A_e) = [f_l(\tilde{A}), f_r(\tilde{A})], \tag{35}$$

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

$$f_l(\tilde{A}) = \min_{\forall A_e} f(A_e) \tag{36}$$

$$f_r(\tilde{A}) = \max_{\forall A_e} f(A_e) \tag{37}$$

$$f(A_e) = 1 - \frac{1}{N}\sum_{i=1}^{N}|2\mu_A(x_i) - 1)|. \quad \square \tag{38}$$

Theorem 2: [29] Let A_{e1} be defined as

$$\mu_{A_{e1}}(x) = \begin{cases} \overline{\mu}_{\tilde{A}}(x), & \overline{\mu}_{\tilde{A}}(x) \text{ is further away from 0.5 than } \underline{\mu}_{\tilde{A}}(x) \\ \underline{\mu}_{\tilde{A}}(x), & \text{otherwise} \end{cases} \tag{39}$$

and A_{e2} be defined as

$$\mu_{A_{e2}}(x)=\begin{cases}\overline{\mu}_{\tilde{A}}(x), & \text{both } \overline{\mu}_{\tilde{A}}(x) \text{ and } \underline{\mu}_{\tilde{A}}(x) \text{ are below } 0.5\\ \underline{\mu}_{\tilde{A}}(x), & \text{both } \overline{\mu}_{\tilde{A}}(x) \text{ and } \underline{\mu}_{\tilde{A}}(x) \text{ are above } 0.5\\ 0.5, & \text{otherwise}\end{cases} \tag{40}$$

Then (36) and (37) can be computed as

$$f_l(\tilde{A})=f(A_{e1}) \tag{41}$$

$$f_r(\tilde{A})=f(A_{e2}). \quad \square \tag{42}$$

C. 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 4: [29] The relative variance of an embedded T1 FS A_e to an IT2 FS $\tilde{A}$, $v_{\tilde{A}}(A_e)$, is defined as

$$v_{\tilde{A}}(A_e)=\frac{\sum_{i=1}^{N}\left[x_i-c(\tilde{A})\right]^2\mu_{A_e}(x_i)}{\sum_{i=1}^{N}\mu_{A_e}(x_i)}, \tag{43}$$

where $c(\tilde{A})$ is the average centroid of $\tilde{A}$ [see (10)]. $\square$

Definition 5: [29] The variance of an IT2 FS $\tilde{A}$, $V(\tilde{A})$, is the union of relative variance of all its embedded T1 FSs A_e, i.e.,

$$V(\tilde{A})\equiv\bigcup_{\forall A_e} v_{\tilde{A}}(A_e)=[v_l(\tilde{A}),v_r(\tilde{A})], \tag{44}$$

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

$$v_l(\tilde{A})=\min_{\forall A_e} v_{\tilde{A}}(A_e) \tag{45}$$

$$v_r(\tilde{A})=\max_{\forall A_e} v_{\tilde{A}}(A_e). \quad \square \tag{46}$$

D. 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.

Definition 6: [29] The relative skewness of an embedded T1 FS A_e to an IT2 FS $\tilde{A}$, $s_{\tilde{A}}(A_e)$, is defined as

$$s_{\tilde{A}}(A_e)=\frac{\sum_{i=1}^{N}[x_i-c(\tilde{A})]^3\mu_{A_e}(x_i)}{\sum_{i=1}^{N}\mu_{A_e}(x_i)}, \tag{47}$$

where $c(\tilde{A})$ is the average centroid of $\tilde{A}$ [see (10)]. $\square$

Definition 7: [29] The skewness of an IT2 FS $\tilde{A}$, $S(\tilde{A})$, is the union of relative skewness of all its embedded T1 FSs A_e, i.e.,

$$S(\tilde{A})\equiv\bigcup_{\forall A_e} s_{\tilde{A}}(A_e)=[s_l(\tilde{A}),s_r(\tilde{A})], \tag{48}$$

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

$$s_l(\tilde{A})=\min_{\forall A_e} s_{\tilde{A}}(A_e) \tag{49}$$

$$s_r(\tilde{A})=\max_{\forall A_e} s_{\tilde{A}}(A_e). \quad \square \tag{50}$$

E. Comparative Studies

Define the average centroid $c(\tilde{A})$ in (10), average cardinality $p(\tilde{A})$ in (34), and

$$f(\tilde{A}) = \frac{f_r(\tilde{A}) + f_l(\tilde{A})}{2} \tag{51}$$

$$v(\tilde{A}) = \frac{v_r(\tilde{A}) + v_l(\tilde{A})}{2} \tag{52}$$

$$|s(\tilde{A})| = \frac{|s_r(\tilde{A})| + |s_l(\tilde{A})|}{2} \tag{53}$$

$$\delta_c(\tilde{A}) = c_r(\tilde{A}) - c_l(\tilde{A}) \tag{54}$$

$$\delta_p(\tilde{A}) = p_r(\tilde{A}) - p_l(\tilde{A}) \tag{55}$$

$$\delta_f(\tilde{A}) = f_r(\tilde{A}) - f_l(\tilde{A}) \tag{56}$$

$$\delta_v(\tilde{A}) = v_r(\tilde{A}) - v_l(\tilde{A}) \tag{57}$$

$$\delta_s(\tilde{A}) = s_r(\tilde{A}) - s_l(\tilde{A}) \tag{58}$$

Then, these quantities can be classified into three groups as follows:

1) Intra-personal uncertainty measures: $p(\tilde{A})$, $f(\tilde{A})$, $v(\tilde{A})$ and $|s(\tilde{A})|$.
2) Inter-personal uncertainty measures: $\delta_c(\tilde{A})$, $\delta_p(\tilde{A})$, $\delta_f(\tilde{A})$, $\delta_v(\tilde{A})$ and $\delta_s(\tilde{A})$.
3) $c(\tilde{A})$ in (10) indicates how large a word is, but it is not an uncertainty measure.

The above ten quantities can also be classified into four groups as follows:

1) μ domain uncertainty measures: $p(\tilde{A})$, $f(\tilde{A})$, $\delta_p(\tilde{A})$ and $\delta_f(\tilde{A})$ [because the summations in (28) and (38) only account for the memberships].
2) x domain uncertainty measures: $v(\tilde{A})$, $|s(\tilde{A})|$, $\delta_v(\tilde{A})$ and $\delta_s(\tilde{A})$ [because the summation in (43) and (47) emphasize the primary variable x].
3) Balanced uncertainty measure: $\delta_c(\tilde{A})$ [because the summation in (7) emphasizes equally on x and $\mu(x)$].
4) Not an uncertainty measure: $c(\tilde{A})$.

The correlations between any two of the above quantities, e.g., $f(\tilde{A})$ and $v(\tilde{A})$, can be computed as

$$correlation(f(\tilde{A}), v(\tilde{A})) = \frac{\sum_{i=1}^{32} f(\tilde{A}_i)v(\tilde{A}_i)}{\sqrt{[\sum_{i=1}^{32} f^2(\tilde{A}_i)][\sum_{j=1}^{32} v^2(\tilde{A}_j)]}} \tag{59}$$

The areas of the 32 word FOUs, as well as the five uncertainty measures, are shown in Table VII. The correlations among the above 10 quantities and the areas of the FOUs are shown in Table VIII. Observe that

1) Except $c(\tilde{A})$, all other nine quantities have strong correlation with the area of the FOU. This is because all other nine quantities are uncertainty measures. As the area of the FOU increases, both intra-personal uncertainty and inter-personal uncertainty increase.
2) Generally $c(\tilde{A})$ has weak correlation with all other quantities because it is not an uncertainty measure.
3) $v(\tilde{A})$ has strong correlation with $|s(\tilde{A})|$, and $\delta_v(\tilde{A})$ has strong correlation with $\delta_s(\tilde{A})$. This is because $V(\tilde{A})$ and $S(\tilde{A})$ measure the second and third order uncertainties in the x domain, whereas all other quantities measure the first order uncertainties.
4) $\delta_c(\tilde{A})$ has strong correlation with all other uncertainty measures.

In summary, the centroid is a very important characteristics for an IT2 FS: its average value can be used in ranking, and its length is a good uncertainty measure.

TABLE VII

UNCERTAINTY MEASURES FOR THE 32 WORD FOUS.

Word	Area of FOU	$C(\tilde{A})$	$P(\tilde{A})$	$F(\tilde{A})$	$V(\tilde{A})$	$S(\tilde{A})$
Teeny-weeny	0.99	[0.05, 1.08]	[0.06, 1.05]	[0 , 0.75]	[0.06, 0.74]	[- 0.14, 0.62]
A smidgen	1.11	[0.21, 1.05]	[0.33, 1.44]	[0.02, 0.70]	[0.10, 0.84]	[- 0.10, 0.96]
None to very little	1.20	[0.13, 1.17]	[0.20, 1.40]	[0.01, 0.72]	[0.10, 1.03]	[- 0.15, 1.18]
Tiny	1.81	[0.21, 1.73]	[0.33, 2.13]	[0 , 0.72]	[0.23, 2.23]	[- 0.48, 3.80]
Very little	1.40	[0.45, 1.49]	[0.11, 1.51]	[0 , 0.83]	[0.03, 0.60]	[- 0.21, 0.54]
Very small	1.10	[0.66, 1.39]	[0.21, 1.31]	[0.04, 0.80]	[0.04, 0.41]	[- 0.10, 0.32]
A bit	1.14	[1.42, 2.09]	[0.52, 1.66]	[0.09, 0.75]	[0.09, 0.52]	[- 0.17, 0.43]
Low amount	2.70	[1.07, 3.13]	[0.19, 2.89]	[0 , 0.82]	[0.06, 2.17]	[- 2.03, 3.67]
Little	2.32	[1.31, 2.95]	[0.30, 2.62]	[0.02, 0.81]	[0.10, 1.73]	[- 1.03, 2.77]
Small	2.81	[1.29, 3.34]	[0.21, 3.02]	[0 , 0.83]	[0.03, 2.06]	[- 2.67, 2.85]
Somewhat small	3.09	[1.73, 4.16]	[0.17, 3.27]	[0 , 0.82]	[0.03, 2.75]	[- 3.58, 4.86]
Some	4.85	[2.03, 5.90]	[0.23, 5.08]	[0 , 0.83]	[0.09, 6.71]	[-13.33, 18.59]
Some to moderate	4.33	[2.75, 6.31]	[0.20, 4.54]	[0 , 0.82]	[0.03, 5.61]	[-12.80, 11.96]
Moderate amount	3.41	[3.50, 6.28]	[0.17, 3.58]	[0 , 0.81]	[0.03, 3.51]	[- 5.54, 6.69]
Fair amount	4.16	[3.47, 6.78]	[0.25, 4.41]	[0 , 0.81]	[0.06, 5.24]	[-12.30, 9.85]
Medium	2.00	[4.19, 6.19]	[0.04, 2.04]	[0 , 0.80]	[0.01, 1.52]	[- 1.56, 1.91]
Modest amount	2.35	[4.57, 6.24]	[0.19, 2.54]	[0 , 0.83]	[0.03, 1.44]	[- 1.36, 1.81]
Good amount	4.05	[5.04, 8.39]	[0.17, 4.22]	[0 , 0.82]	[0.10, 5.04]	[-12.61, 7.59]
Sizeable	2.92	[6.16, 8.15]	[0.34, 3.27]	[0 , 0.82]	[0.10, 2.31]	[- 4.04, 2.24]
Quite a bit	2.92	[6.16, 8.15]	[0.34, 3.27]	[0 , 0.82]	[0.10, 2.31]	[- 4.04, 2.24]
Considerable amount	3.31	[5.97, 8.52]	[0.19, 3.50]	[0 , 0.83]	[0.06, 3.10]	[- 6.10, 3.64]
A lot	2.59	[6.99, 8.82]	[0.27, 2.85]	[0 , 0.82]	[0.07, 1.92]	[- 3.15, 1.54]
High amount	2.47	[7.19, 8.82]	[0.38, 2.84]	[0.02, 0.82]	[0.13, 1.83]	[- 3.01, 1.08]
Substantial amount	2.47	[7.19, 8.82]	[0.38, 2.84]	[0.02, 0.82]	[0.13, 1.83]	[- 3.01, 1.08]
Very sizeable	2.79	[6.94, 9.10]	[0.17, 2.97]	[0 , 0.83]	[0.13, 2.51]	[- 4.51, 1.70]
Large	1.87	[7.50, 8.75]	[0.32, 2.19]	[0.04, 0.80]	[0.10, 1.18]	[- 1.55, 0.47]
Very large	0.92	[9.03, 9.57]	[0.68, 1.61]	[0.06, 0.66]	[0.12, 0.57]	[- 0.55, 0.06]
Humongous amount	1.27	[8.70, 9.91]	[0.13, 1.40]	[0 , 0.73]	[0.10, 1.18]	[- 1.33, 0.23]
Huge amount	1.24	[8.88, 9.82]	[0.28, 1.51]	[0 , 0.73]	[0.10, 0.93]	[- 1.06, 0.12]
Very high amount	1.06	[8.97, 9.78]	[0.35, 1.40]	[0.03, 0.69]	[0.10, 0.81]	[- 0.92, 0.09]
Extreme amount	0	[9.56, 9.56]	[0.70, 0.70]	[0.46, 0.47]	[0.10, 0.10]	[- 0.02,- 0.02]
Maximum amount	0.49	[9.50, 9.87]	[0.21, 0.70]	[0.04, 0.67]	[0.03, 0.18]	[- 0.10, 0.01]

TABLE VIII

CORRELATIONS AMONG DIFFERENT UNCERTAINTY MEASURES.

	Area	$c(\tilde{A})$	$p(\tilde{A})$	$f(\tilde{A})$	$v(\tilde{A})$	$\|s(\tilde{A})\|$	$\delta_c(\tilde{A})$	$\delta_p(\tilde{A})$	$\delta_f(\tilde{A})$	$\delta_v(\tilde{A})$	$\delta_s(\tilde{A})$
Area	1	.75	.99	.89	.97	.88	1	1	.92	.96	.88
$c(\tilde{A})$	.75	1	.80	.86	.67	.52	.73	.75	.82	.64	.52
$p(\tilde{A})$	.99	.80	1	.94	.95	.84	.98	.99	.94	.93	.84
$f(\tilde{A})$	.89	.86	.94	1	.80	.64	.88	.89	.98	.78	.64
$v(\tilde{A})$	.97	.67	.95	.80	1	.96	.98	.97	.83	1	.96
$s(\tilde{A})$	.88	.52	.84	.64	.96	1	.89	.88	.67	.97	1
$\delta_c(\tilde{A})$	1	.73	.98	.88	.98	.89	1	1	.91	.97	.89
$\delta_p(\tilde{A})$	1	.75	.99	.89	.97	.88	1	1	.92	.96	.88
$\delta_f(\tilde{A})$	.92	.82	.94	.98	.83	.67	.91	.92	1	.81	.67
$\delta_v(\tilde{A})$	.96	.64	.93	.78	1	.97	.97	.96	.81	1	.97
$\delta_s(\tilde{A})$	.88	.52	.84	.64	.96	1	.89	.88	.67	.97	1

VI. Connections to Psychological Rules

From psychologists' view, Wallsten and Budescu [23] pointed out that *"when combining, comparing, or trading-off information about uncertainty with information about other dimensions, such as outcome values, the uncertainty representation, $\mu(x)$, is converted from a vague interval to a point value by restricting attention only to values of x with membership above a threshold, θ, i.e., for which $\mu(x) \geq \theta$. A specific value, p^*, is then selected probabilistically according to a weighting function proportional to the $\mu(x) \geq \theta$."* Let $\theta = 0$, and the weighting function be

$$\frac{\sum_{i=1}^N x_i \mu(x_i)}{\sum_{i=1}^N \mu(x_i)}. \tag{60}$$

Then, (60) is actually the centroid of a T1 FS. This justifies why the (average) centroid can be used in ranking FSs.

Wallsten and Budescu [23] also pointed out that *"when combing separate components of information about a single dimension, the resulting judgement reflects some sort of average of their values."* Next we shall show that Perceptual Reasoning (PR) [18], [30], a CWW engine using a special LWA ($\tilde{W}_i$ in (1) are interval firing levels of the rules instead of words modeled by IT2 FSs), satisfies this rules.

In Perceptual Reasoning, we consider the following problem [30]:

Given a rulebase with N rules, each of the form:

$$R^i: \textit{ If } x_1 \textit{ is } \tilde{F}_1^i \textit{ and } \ldots \textit{ and } x_p \textit{ is } \tilde{F}_p^i, \textit{ Then } y \textit{ is } \tilde{G}^i \tag{61}$$

where $\tilde{F}_j^i$ and $\tilde{G}^i$ are words modeled by IT2 FSs (the words and their FOUs constitute a codebook), and a new input

$$\tilde{\mathbf{X}}' = (\tilde{X}_1, \ldots, \tilde{X}_p), \tag{62}$$

where $\tilde{X}_i$ are also words (from the codebook) modeled by IT2 FSs, then what is the output IT2 FS $\tilde{Y}_{PR}$ and its associated word in the codebook?

It has been shown [30] that "*for PR using IT2 FSs, $\tilde{Y}_{PR}$ cannot be smaller than the smallest consequent of the fired rules, and it also cannot be larger than the largest consequent of the fired rules,*" i.e., $\tilde{Y}_{PR}$ is some sort of average of the consequents of the fired rules. A graphical illustration of this property is shown in Fig. 5. Assume only two rules are fired and $\tilde{G}^1$ lies to the left of $\tilde{G}^2$; then, $\tilde{Y}_{PR}$ lies between $\tilde{G}^1$ and $\tilde{G}^2$.

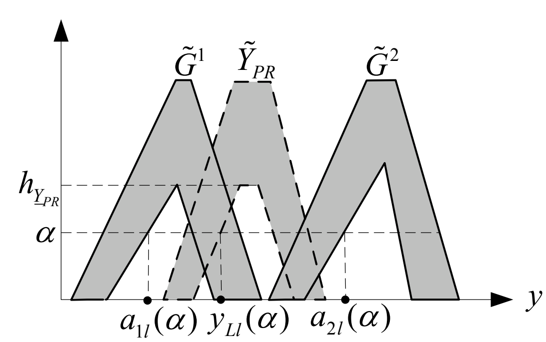

Fig. 5. A graphical illustration of PR, when only two rules fire.

VII. Conclusions

In this report, several ranking methods, similarity measures and uncertainty measures for IT2 FSs were evaluated based on real survey data, and a centroid-based ranking method and the LWA were tested against psychological rules. It is shown that:

1) The new proposed centroid-based ranking method is better than Mitchell's ranking method for IT2 FSs.
2) Both the vector similarity measure and the new proposed Jaccard's similarity measure are better than all other similarity measures for IT2 FSs; however, Jaccard's similarity measure has lower computational cost, and hence it is more favorable in practice.
3) The centroid of an IT2 FS is more important than other uncertainty measures: the average centroid can be used in ranking, and the spread of the centroid is a well-balanced uncertainty measure.
4) The centroid-based ranking method and the LWA coincide with psychological rules.

The research results will help people better understand the uncertainties associated with linguistic terms and hence how to use them effectively in survey design and linguistic information processing.

Acknowledgment

This work is supported by the IEEE Computational Intelligence Society Walter Karplus Summer Research Grant. The author would like to thank Professor Jerry M. Mendel with the University of Southern California and Professor David V. Budescu with the University of Illinois at Urbana-Champaign for their support and very helpful comments.

References

[1] [Online]. Available: http://sipi.usc.edu/ mendel/software/

[2] H. Bustince, "Indicator of inclusion grade for interval-valued fuzzy sets. application to approximate reasoning based on interval-valued fuzzy sets," *Int'l Journal of Approximate Reasoning*, vol. 23, no. 3, pp. 137–209, March 2000.

[3] A. De Luca and S. Termini, "A definition of nonprobabilistic entropy in the setting of fuzzy sets theory," *Information and Computation*, vol. 20, pp. 301–312, 1972.

[4] M. B. Gorzalczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 21, pp. 1–17, 1987.

[5] D. Harmanec, "Measures of uncertainty and information," 1999, http://www.sipta.org/documentation/summary_measures/main.html.

[6] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bulletin de la Societe de Vaud des Sciences Naturelles*, vol. 44, p. 223, 1908.

[7] N. N. Karnik and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences*, vol. 132, pp. 195–220, 2001.

[8] G. J. Klir, "Principles of uncertainty: What are they? why do we need them?" *Fuzzy Sets and Systems*, vol. 74, pp. 15–31, 1995.

[9] F. Liu and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," submitted to IEEE Trans. on Fuzzy Systems, 2007.

[10] J. M. Mendel, "Computing with words, when words can mean different things to different people," in *Proc. 3rd Int'l ICSC Symposium on Fuzzy Logic and Applications*, Rochester, NY, June 1999, pp. 158–164.

[11] ——, "An architecture for making judgment using computing with words," *Int'l Journal of Applied Mathematics and Computer Science*, vol. 12, no. 3, pp. 325–335, 2002.

[12] ——, "Computing with words and its relationships with fuzzistics," *Information Sciences*, vol. 177, pp. 988–1006, 2007.

[13] J. M. Mendel and D. Wu, "Signal fusion using novel weighted averages," submitted to *Information Fusion*, 2007.

[14] J. M. Mendel and H. Wu, "Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 1, forward problems," *IEEE Trans. on Fuzzy Systems*, vol. 14, no. 6, pp. 781–792, 2006.

[15] ——, "New results about the centroid of an interval type-2 fuzzy set, including the centroid of a fuzzy granule," *Information Sciences*, vol. 177, pp. 360–377, 2007.

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

[17] J. M. Mendel and D. Wu, “Perceptual reasoning: A new computing with words engine,” in *Proc. IEEE Int’l Conf. on Granular Computing*, Silicon Valley, November 2007, pp. 446–451.

[18] ——, “Perceptual reasoning for perceptual computing,” submitted to *IEEE Trans. on Fuzzy systems*, June 2007.

[19] J. M. Mendel and H. Wu, “Type-2 fuzzistics for non-symmetric interval type-2 fuzzy sets: forward problems,” *IEEE Trans. on Fuzzy Systems*, 2007, , In Press.

[20] H. B. Mitchell, “Pattern recognition using type-II fuzzy sets,” *Information Sciences*, vol. 170, pp. 409–418, 2005.

[21] ——, “Ranking type-2 fuzzy numbers,” *IEEE Trans. on Fuzzy Systems*, vol. 14, no. 2, pp. 287–294, 2006.

[22] M. Nikravesh, “Soft computing for reservoir characterization and management,” in *Proc. IEEE Int’l Conf. Granular Computing*, vol. 2, Beijing, China, July 2005, pp. 593–598.

[23] T. S. Wallsten and D. V. Budescu, “A review of human linguistic probability processing: General principles and empirical evidence,” *The Knowledge Engineering Review*, vol. 10, no. 1, pp. 43–62, 1995.

[24] X. Wang and E. E. Kerre, “Reasonable properties for the ordering of fuzzy quantities (I),” *Fuzzy Sets and Systems*, vol. 118, pp. 375–387, 2001.

[25] ——, “Reasonable properties for the ordering of fuzzy quantities (II),” *Fuzzy Sets and Systems*, vol. 118, pp. 387–405, 2001.

[26] D. Wu and J. M. Mendel, “The linguistic weighted average,” in *Proc. FUZZ-IEEE*, Vancouver, BC, Canada, July 2006, pp. 566–573.

[27] ——, “Aggregation using the linguistic weighted average and interval type-2 fuzzy sets,” *IEEE Trans. on Fuzzy Systems*, 2007, in press.

[28] ——, “Enhanced Karnik-Mendel Algorithms,” *IEEE Trans. on Fuzzy Systems*, 2007, in press.

[29] ——, “Uncertainty measures for interval type-2 fuzzy sets,” *Information Sciences*, vol. 177, no. 23, pp. 5378–5393, 2007.

[30] ——, “Perceptual reasoning using interval type-2 fuzzy sets: Properties,” in *Proc. IEEE World Congress on Computational Intelligence*, Hong Kong, June 2008.

[31] ——, “A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets,” *Information Sciences*, vol. 178, no. 2, pp. 381–402, 2008.

[32] L. A. Zadeh, “Fuzzy logic = computing with words,” *IEEE Trans. on Fuzzy Systems*, vol. 4, pp. 103–111, 1996.

[33] ——, “From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions,” *IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 4, pp. 105–119, 1999.

[34] W. Zeng and H. Li, “Relationship between similarity measure and entropy of interval valued fuzzy sets,” *Fuzzy Sets and Systems*, vol. 157, pp. 1477–1484, 2006.