DOI: https://doi.org/10.1109/FUZZ-IEEE.2019.8858947

Publisher: IEEE

Version: Accepted Version

Ontological and Fuzzy Set Similarity between Perception-Based Words

Valerie Cross
Computer Science and Software Engineering
Miami University
Oxford, OH USA
crossv@miamioh.edu

Valeria Morenko
Computer Science and Software Engineering
Miami University
Oxford, OH USA
mokrenvi@miamioh.edu

Keeley Crockett
Computational Intelligence Lab
Manchester Metropolitan University
Manchester, UK
K.Crockett@mmu.ac.uk

Naeemeh Adel
School of Computing, Maths, and Digital Technology
Manchester Metropolitan University
Manchester, UK
N.Adel@mmu.ac.uk

***Abstract*—Fuzzy short text semantic similarity measures allow the inclusion of human perception based words to be within the similarity measurement which results in better correlation on the meaning of the short text with human understanding. Existing measures such as FUSE and FAST rely on the creation of fuzzy ontological structures from the modelling of perception words using type-1 or type-2 fuzzy sets. Due to the complex methodology of creating these ontologies, fuzzy word representation cannot be guaranteed due to language evolution. This paper presents a comparative study of simpler fuzzy set similarity measures. The results surprisingly indicate that a very simple fuzzy set similarity measure created from the center of gravity (COG) distance between type-2 fuzzy sets has a very high correlation with the FUSE semantic similarity measure.**

Keywords—ontology, semantic similarity, fuzzy set similarity, human perception

I. INTRODUCTION

A goal of artificial intelligence is to develop machines that communicate and understand natural language. Communication between machines uses crisp quantities, but an important characteristic of natural language is many words are vague or imprecise. Vagueness often exists in domain knowledge as understood by humans. Often humans communicating with each other or providing domain knowledge are more comfortable using inexact, vague terms, or perception-based, that is, fuzzy words that are subjective. For humans and machines to communicate and for machines to understand domain knowledge, a method of interpreting fuzzy words is needed. Computing with Words (CWW) [1] provides the ability to interpret these fuzzy words. Fuzzy set theory and CWW research presents essential concepts necessary to make progress towards the goal of finding representations of natural language or fuzzy words used by humans and reasoning with these representations.

Handlubg uncertainty in human language has motivated the natural language processing research community to develop sentence similarity measures. Early work focused on syntactic similarity [2]. Latent Semantic Analysis [3] brought in semantic similarity between blocks of text by producing statistics based on occurrences of the words in the blocks within a large corpus. Using statistical analysis, LSA creates semantic vectors. It calculates similarity between these vectors. Following this, STASIS [4] examined the use of semantic similarity measures within the context of an ontology, a knowledge structure containing concepts and defining relationships between these concepts. Much research exists on semantic similarity measures, also referred to as ontological similarity measures [5], between concepts in an ontology. For measuring text similarity for short pieces of text, STASIS uses the WordNet ontology and a semantic similarity measure [6] between each word pair, one word from each text, to create a semantic vector and incorporates corpus statistics in the semantic vector. STASIS integrates the early approach of measuring syntactic similarity into the final similarity measure between two pieces of text.

Although this previous research made progress in measuring text similarity, it failed to address the occurrence of imprecise and vague words, i.e., fuzzy words that occur extensively in natural language. This capability is needed in order to advance conversational understanding between humans and machines. Since different people have different interpretations or meanings for fuzzy words, singular quantities for them are not reasonable. Fuzzy sets serve as a means of representing fuzzy words. CWW provides a framework by which fuzzy words can be quantified, scaled against each other and then become machine representable. The scaling of fuzzy words through obtaining human perceptions is a critical step for creating fuzzy sentence similarity measures.

FAST (Fuzzy Algorithm for Similarity Testing) [7] was developed to measure the similarity between pairs of fuzzy words and incorporate this additional similarity evaluation into the overall sentence similarity measure between sentences or pieces of short text. To accomplish this, it was necessary to create a dataset containing quantified fuzzy words which are organized into six different categories [7]: *age, size/distance, frequency, goodness, membership level* and *temperature*. In a comparative experimental study, FAST demonstrated an improvement in measuring semantic sentence similarity over existing algorithms STASIS and LSA, which are unable to process fuzzy words in text.

More recent research developed FUSE (FUzzy Similarity mEasure) [8] which extends the FAST research to address the differences between modeling fuzzy words with type-1 versus type-2 fuzzy sets. In FAST, human experts were used to create type-1 fuzzy sets for the fuzzy words; however, on further consideration, it was felt that these fuzzy sets were not accurate representations because of the subjective nature of the human evaluators. Essentially, type-1 fuzzy sets could not capture the uncertainty of humans [9]. FUSE uses specifically type-2 interval fuzzy sets since they are simpler to use because the membership functions are interval sets. FUSE also has a larger vocabulary across the six categories with over 57% increased coverage of fuzzy words. Both FUSE and FAST, however rely on pre-constructed fuzzy ontologies, resulting in complex measures, which will not perform well if there is not extensive modelling of fuzzy words for any given language.

This paper focuses on the measurement of similarity between fuzzy words represented as type-1 fuzzy sets using three different existing fuzzy set similarity measures. These fuzzy sets are directly created from the data collected from the human evaluators. This approach is simpler than that of FAST and FUSE for measuring fuzzy word similarity. Because type-2 fuzzy sets may better represent the subjective nature of a fuzzy word and are used in FUSE, a fourth similarity measure using a scaled COG for the type-2 fuzzy word representations is also used in our study. The objective is to determine how well these simpler fuzzy set similarity measures correlate with the semantic similarity measure used in FAST and FUSE.

The paper organization is as follows: Section II first examines some of the difficulties when using humans to gather data for the process of defining fuzzy words as fuzzy sets and describes the approaches to representing fuzzy words to measure similarity between them. Section III describes the approach for fuzzy word representation used in this paper's research. It reviews the existing fuzzy set similarity measures and a simple similarity measure calculated from the distance between the COGs for two fuzzy words represented as type-2 interval fuzzy sets. Section IV describes the experimental design and compares the results from applying these measures to word pairs used in previous studies [4] [7] [8]. Finally, Section V presents the conclusions and future work.

II. CONTEXT OF FUZZY WORDS AND THEIR REPRESENTATION

A. Type-1 versus Type-2 Fuzzy Sets

In [1] a fuzzy set (type-1) representation is described as a means of defining perception-based or fuzzy words. Type-2 fuzzy sets [9] were developed to address the issue of perception-based words varying from individual to individual. Instead of using a single fuzzy set, a set of fuzzy sets represents a fuzzy word; that is, a type-2 fuzzy set is a set wherein all its elements are fuzzy type-1 sets. In FAST, type-1 fuzzy sets are developed for fuzzy words but in FUSE type-2 interval fuzzy sets are used. In both of these approaches ontologies are created to represent the relationships among the fuzzy words in six different categories. The six categories are broad enough to hold a large range of fuzzy word and allow related fuzzy words to be scaled in terms of association within the category. These ontologies are created by scaling a representative value of the fuzzy set into the interval [-1, +1]. The scaled value determines into which node of the ontology the fuzzy word is placed.

B. Creating the fuzzy word representation for FAST

Two empirical experiments were undertaken with human subjects. The first required the subjects to populate the six categories with fuzzy words. Next subjects had to quantify the fuzzy words in each category. Quantification was done using a scale of 0 to 10. The subjects were asked to specify a single value, a point in the 0 to 10 scale where the membership function for a fuzzy word would be highest. For each fuzzy word the mean and standard deviation values were calculated from all the subjects' ratings for that fuzzy word. Then the relationships among the fuzzy words within a category were established by creating ontologies based on these values. These ontologies of fuzzy words are needed since the semantic similarity measure used between two fuzzy words is that in [6]. Although numerous semantic similarity measures have been proposed over the years [5], this research focuses on the specific measure used in the FAST and FUSE research which addresses some of the weaknesses of the older semantic similarity measures. The formula for the semantic similarity measure, S used to determine word pair similarity of words, w_1 and w_2 is

$$\boxed{S(w_1, w_2) = e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}}}$$

where l represents the path length between the two words in the ontology and h represents the depth of their common subsumer. For FAST, the parameters α and β were set to 0.2 and 0.6, respectively and were determined empirically.

To create the category ontologies, five nodes were established for each category. The root node for each category contains those fuzzy words whose mean values were around the midpoint value (within the 0 to 10 range). This root node is labeled AVERAGE for each category. As an example, for the *size/distance* category, the five nodes are labeled {VERY SMALL, SMALL AVERAGE, LARGE, VERY LARGE}. Examples of fuzzy words in its root node include *medium* and *middle.* From the root node there are two branches. To the left are two nodes for the fuzzy words with lower mean values, {VERY SMALL, SMALL}. To the right are two nodes for the fuzzy words with higher mean values, {LARGE, VERY LARGE}. To place the fuzzy words in the appropriate nodes, the mean values were re-scaled to a range of -1 to +1 and then a range of re-scaled values was established for each node and used to determine to which node a fuzzy word should be assigned. Each category ontology was created in this manner; for example, the *temperature* category has the nodes {VERY COLD, COLD, AVERAGE, HOT, VERY HOT}. FAST uses the created category ontologies with the semantic similarity measure in [6] to determine the similarity between pairs of fuzzy words. This word pair similarity measurement is one component of the FAST algorithm that establishes a measure of text similarity between pairs of sentences or pieces of text.

C. Creating the fuzzy word representation for FUSE

FUSE takes a similar approach to FAST in that in creates ontologies based on the six categories and the fuzzy words within those categories; however, it expanded on the number of fuzzy words since FAST had only 196 words within the six categories. It did this by taking the existing FAST words and adding only the one word synonyms for these words that could be found in a dictionary. This process resulted in a total of 309 fuzzy words over the six categories.

As in FAST, human subjects are used to construct the fuzzy sets for the fuzzy wordsT. hese fuzzy sets are based on Mendel's Hao-Mendel Approach (HMA) using type-2 interval fuzzy sets [13] to collect data from the subjects. The same 0 to 10 range is kept. The subjects are asked to provide an interval value for the fuzzy word instead of a single value as in FAST. This interval value represents the range where the subject believes the fuzzy word should be placed in the range of 0 to 10. Noise is eliminated by removing bad data and outliers.

From the cleaned up data, the center of gravity (COG) was determined using the upper and lower footprints of uncertainty. As in FAST, the COG value for a fuzzy word was scaled into the -1 to +1 range in order to create the ontology. FUSE, however, increased the number of nodes for a category ontology from 5 to 11 and the root node was an arbitrary category label node. The ontology became a binary tree with nodes containing negative values on the left side of the root node and nodes containing positive values on the right side. The fuzzy words were grouped using a 0.2 interval size. As in FAST, the similarity measure given in [6] was used with these category ontologies to determine semantic similarity between pairs of fuzzy words. The parameters α and β for FUSE were determined empirically and set to 0.15 and 0.85, respectively.

III. Fuzzy Set Similarity Measure Between Fuzzy Words

The approaches to measuring fuzzy word similarity in STASIS, FAST and FUSE have as their basis semantic or ontological similarity measures within an ontology structure. A detailed review of semantic similarity measures can be found in [5]. The FAST and FUSE approaches require creating ontologies for each of the six categories so that a semantic similarity measure can be used between the fuzzy words. The approach used in our research does not require creating ontologies, Instead three fuzzy set similarity measures are used between triangular fuzzy sets created from the FAST type-1 fuzzy sets. The fourth similarity measure uses the distance between the normalized centers of gravity (COG) for type-2 interval fuzzy sets created for FUSE.

A. Creation of Trianglar Fuzzy Sets

For purposes of the FAST experiments data from the type-1 fuzzy sets were acquired from the FAST researchers, specifically the defuzzified value or mean and the standard deviation. With these values, a pseudo triangular fuzzy set is created where the membership degree at the mean value is 1.0. A normal probability density distribution is used and values ±3 standard deviations away from the mean were used for the end points of the triangular fuzzy set since 99.7% of the data is within three standard deviations of the mean. See Fig. 1 that shows the triangular membership function for *centre* with a mean of 4.93 and a standard deviation of 0.5. The simplest approach to building fuzzy sets for fuzz words is used since the hypothesis is to determine if these sets based on human judgment might be used with well-known fuzzy set similarity measures to eliminate the need to build ontologies.

Twenty word pairs selected from those in [7] are used to compare measures. Triangular fuzzy sets are created for each fuzzy word. Fuzzy set similarity measures can simply be used between the triangular membership functions. This approach is more efficient since the category ontologies creation is eliminated. Experiments are described in the following section with the specific fuzzy set similarity measures discussed here. The fuzzy word pair similarities are produced to determine how closely the results correlate with those produced by STASIS, FAST and FUSE; all of which use the same semantic similarity measure within an ontology.

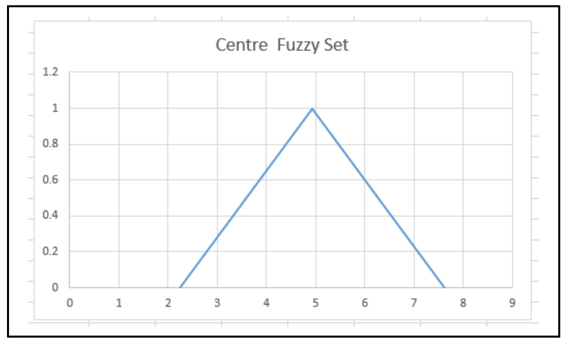

Fig. 1 Centre fuzzy set

The first three fuzzy set similarity measures described are used on the triangular membership functions. The last one uses the COGs of the type-2 interval fuzzy sets. Zadeh's sup-min is a partial matching measure [10]. The fuzzy Jaccard is a fuzzy set equality measure [11]. GeoSim uses the geometric distance between fuzzy sets to determine similarity [12]. The COG similarity measure for type-2 simply takes the distance between normalized COGs for the two fuzzy sets. It then normalizes the distance and converts it to similarity by subtracting from one. Both sup-min and the Jaccard measures produce a 0 similarity when the two fuzzy sets do not overlap. GeoSim and the COG type-2 similarity measures, however, produce a non-zero value even when the fuzzy sets do not overlap since both are based on distance.

B. Sup-Min

In [10] a detailed and thorough review of a variety of fuzzy set similarity measures is provided. Zadeh's consistency index also known as the sup-min or partial matching index falls into the set-theoretic category of fuzzy similarity measures. It roughly estimates the similarity between two fuzzy sets by finding at what domain values they intersect and determines their similarity by taking the highest membership degree among their intersection points. Given two fuzzy sets A and A', similarity between the two is determined as

$$S_{Zadeh}(A, A') = \sup_{u \in U} T(A'(u), A(u)) \quad (1)$$

where T can be any t-norm, but usually the minimum is used for the t-norm. It is referred to as a partial since it only provides an estimated similarity value between the two fuzzy sets. .

C. Jaccard

The fuzzy Jaccard similarity measure is defined as a fuzzy extension of the Jaccard index [11] between two crisp sets by replacing set cardinality with fuzzy set cardinality. This fuzzy set similarity measure is also in the set theoretic category but provides a more comprehensive view of similarity between the two fuzzy sets since all elements in both fuzzy sets are taken into account not just the intersection point as in sup-min. Given two fuzzy sets A and A', similarity between the two is determined as

$$S_{Jaccard}(A, A') = | A \cap A'| \;/\; | A \cup A'| \quad (2)$$

so the similarity is measured by the proportion of the area of the intersection of the two fuzzy sets to the area of the union of the two fuzzy sets.

D. Geometric Fuzzy Similarity Based on Dissemblance Index

Set theoretic fuzzy set similarity measures do not consider the distance of the fuzzy set A' from A. With the geometric fuzzy similarity measure [12], the distance between the two sets is the basis for determining their similarity. This distance is based on the dissemblance index that measures the distance between two real intervals. If $V = [v_1, v_2]$ and $W = [w_1, w_2]$, then

$$DI(V,W) = (|v_1 - w_1| + |v_2 - w_2|) / [2(\beta_2 - \beta_1)] \quad (3)$$

where $[\beta_1, \beta_2]$ is an interval that contains both V and W. The factor $2(\beta_2 - \beta_1)$ is necessary to produce a normalized degree of dissemblance such that $0 \leq D(V, W) \leq 1$. The dissemblance index consists of two components, the left and right sides of each interval and may be generalized to fuzzy intervals.

A fuzzy interval N is defined by a pair of boundary functions L and R and parameters $(r_1, r_2, \lambda, \rho)$. The core of N, the values for which $\mu_N(r) = 1.0$ is the interval $[r_1, r_2]$. Parameters λ and ρ are used to define the left L and the right R boundary functions and the support of N, the values for which $\mu_N(r) \geq 0$, which is $[r_1 - \lambda, r_2 + \rho]$. The L function and the R function define the membership functions for elements in the intervals $[r_1 - \lambda \;\; r_1]$ and $[r_2, r_2 + \rho]$, respectively. If L is positively sloping and linear and R is negatively sloping and linear then the interval N is a trapezoidal fuzzy membership function. Calculating the fuzzy dissemblance index between A and A' is done as an integration over α in the range 0 to 1 as

$$fDI(A'(u),A(u)) = [\textstyle\int |L_{A'}(\alpha) - L_A(\alpha)| + |R_{A'}(\alpha) - R_A(\alpha)| d\alpha] / [2(\beta_2 - \beta_1)] \quad (4)$$

where $[\beta_1, \beta_2]$ is an interval that contains both A' and A. fDI calculates a dissimilarity measure between the two fuzzy intervals based on a normalized distance. It can be converted into a similarity measure between the fuzzy intervals as

$$S_{GeoSim}(A, A') = 1 - fDI(A(u), A'(u)) \quad (5)$$

With this similarity measure, even though A and A' may not overlap, a nonzero similarity value is produced since distance between the two sets is used.

E. Similarity on Type-2 Defuzzified Values Distance

As previously explained in [8] type-2 interval fuzzy sets were used and then defuzzified into a single value by adapting Mendel's footprint of uncertainty (FOU) method [13]. For each word in the six categories, the COG was determined using the lower FOU and upper FOU. The COGs were then scaled into the range [-1, +1]. To see how well a measure based solely on the distance between these scaled COG values worked, the following simple similarity measure is also used in this study:

$$S_{Type2\text{-}Dist}(A, A') = 1 - | COG_{Scaled}(A) - COG_{Scaled}(A')| / 2 \quad (6)$$

The distance between the two centers of gravity is normalized by the size of the scaled interval [-1, +1]. Calculating this similarity measure between pairs of fuzzy words provides a means of determining how well it correlates with the ontology-based similarity measure developed for FAST and FUSE.

IV. Experimental Results and Analysis

Table I shows 20 fuzzy word pairs used in the experimental investigation. These pairs were taken from the 30 sentence pairs used in the FAST study on sentence similarity [7]. Each of the sentences in the 30 pairs contain only one fuzzy word. Only 20 fuzzy word pairs are selected since 10 pairs are not both from the same category. Although a limited number of pairs, they can still provide evidence of proof of concept for the use of fuzzy set similarity measures. Once more data becomes available, more experiments can be undertaken.

Table I shows the similarity values produced by the various measures. STASIS, FAST and FUSE similarity values are all determined using the semantic similarity measure in [6] and differ because they use different ontological structures. STASIS uses WordNet. FAST uses the fuzzy category ontologies, each having five nodes in a binary tree structure and derived from the type-1 fuzzy sets created for each fuzzy word. FUSE also uses category ontologies; however, each has 11 nodes with a binary tree structure with 5 nodes on each side of the tree. Type-2 interval fuzzy sets are used to derive the FUSE category ontologies.

The correlations between the various pairs of similarity measures are presented in Table II. One can clearly see that STASIS has the lowest correlation with all the other similarity measures. That is an expected result since STASIS does not handle fuzzy words but uses the semantic similarity measure in [6] with the WordNet ontology. Its highest correlations are with FAST at over 0.46 and with FUSE at almost 0.39. Both of these use the same semantic similarity measure as STASIS, however, they use their own ontology categories instead of WordNet. The higher correlation of STASIS with FAST is most likely due to the FAST's simpler ontological structure so that the effects of fuzzy word similarity measure is not as significant as that for FUSE.

TABLE I. SIMILARITY VALUES

Pair	Word1	Word2	GeoSim	Zadeh	Jaccard	Type2-Dist	STASIS	FAST	FUSE
WP1	Short	Massive	0.580804	0.286267	0.042663	0.363095	0.150000002	0.15	0.535246
WP5	Large	Small	0.675234	0.524374	0.159339	0.404762	0.932427814	0.932428	0.932428
WP7	Young	Youthful	0.869565	0.851666	0.568999	0.963768	0.927062972	0.998243	0.99972
WP8	Tiny	Large	0.60219	0.346733	0.063895	0.479167	0.150000002	0.581294	0.616686
WP9	Always	Always	1	1	1	1	1	1	1
WP10	Long	Mammoth	0.849206	0.824403	0.514693	0.895105	0.150000002	0.9857	0.9857
WP11	Minuscule	Enormous	0.497379	0	0	0.142858	0.150000003	0.15	0.384909
WP13	Tiny	Diminutive	0.841967	0.976682	0.400187	0.988096	0.949173456	0.999965	0.999741
WP15	Midpoint	Centre	0.887037	0.995291	0.555488	1	0.999992462	0.999992	0.999992
WP16	Lukewarm	Hot	0.6525	0.473349	0.126103	0.806229	0.150000004	0.831912	0.922408
WP18	Good	Great	0.837662	0.808354	0.485216	0.968966	0.528628322	0.956196	0.99901
WP19	Huge	Small	0.59866	0.337049	0.06016	0.232143	0.150000002	0.584372	0.428318
WP20	Immense	Great	0.9125	0.927778	0.723266	0.953164	0.150000002	0.999984	0.999984
WP22	Great	Long	0.896104	0.885346	0.644481	0.988252	0.150000002	0.999982	0.999982
wp23	Nearby	Faraway	0.53271	0.132552	0.008801	0.491072	0.150000004	0.15	0.620278
WP24	Great	Small	0.598152	0.335649	0.059631	0.315477	0.392989907	0.57195	0.475895
WP25	Loads	Gargantuan	0.850622	0.919384	0.510272	0.666667	0.150000002	0.817206	0.817206
WP27	Excellent	Wonderful	0.881061	0.958561	0.584655	0.87931	0.150000004	1	0.982184
WP29	Large	Oversized	0.926075	0.99689	0.709608	0.863879	0.150000002	0.15	0.98051
WP30	Big	Massive	0.871843	0.87019	0.598059	0.955357	0.150000004	0.925136	1

Removing STASIS from the comparison since it does not handle fuzzy words, FUSE has the highest correlation with all the other similarity measures. Note that its correlations for all the fuzzy set similarity measures are greater than 0.80 and so greater than its correlation of about 0.74 with FAST. FAST is basically a precursor to FUSE with the noted differences for FUSE of type-2 interval fuzzy sets versus type-1 in FAST and the more complex 11 node category ontology versus only the 5 node category ontology in FAST.

It is surprising to see the simple fuzzy set similarity measure $S_{Type2\text{-}Dist}$ has the highest correlation 0.931708 with the more complex FUSE since it requires building ontologies for each of the six categories and using semantic similarity within an ontology. The $S_{Type2\text{-}Dist}$ simply takes the distance between the normalized COGs for the two fuzzy words, normalizes that distance based on the [-1, +1] interval, and converts it to a fuzzy similarity measure by subtracting it from 1.

TABLE II. CORRELATIONS BETWEEN SIMILARITY VALUES

	STASIS	FAST	FUSE
GeoSim	0.331881	0.673149	0.874064
Zadeh	0.354013	0.70461	0.88457
Jaccard	0.286197	0.585729	0.804873
Type2-Dis	0.316763	0.693164	0.931708
STASIS	1	0.461274	0.387813
FAST	0.461274	1	0.736067
FUSE	0.387813	0.736067	1

Table III shows summary statistics for the similarity measures given in Table I.

TABLE III. SUMMARY STATISTICS FOR SIMILARITY VALUES

	GeoSim	Zadeh	Jaccard	Type2-Dist	STASIS	FAST	FUSE
Averages	0.768064	0.672526	0.390776	0.717868	0.384014	0.739218	0.83400984
Std Dev	0.155344	0.329423	0.298685	0.298296	0.355915	0.3352224	0.22720339
Low	0.497	0	0	0.142858	0.15	0.15	0.38490874
Low WP	11	11	11	11	13 pairs	1, 11,23,29	11
High	1	1	1	1	1, 0.99999	1, 0.99999	1, 0.99999
High WP	9	9	9	9, 15	9,15	9,27, 15	9,30, 15

As can be seen in Table I, all similarity measures agree on at least one word pair with the smallest similarity value, that is, word pair 11. However, only the Zadeh and Jaccard measures return 0 for this pair since there is no overlap between the triangular membership functions for those two fuzzy words. STASIS produces 0.15 similarity for 13 of the 20 word pairs and FAST produces 0.15 similarity for 4 of the 20 pairs and agrees with STASIS on those same 4 pairs. Since STASIS cannot handle fuzzy words, it can only use the semantic similarity measure as applied within the WordNet ontology and, therefore, cannot discriminate between these 13 pairs. FAST improves upon STASIS but still produces 4 pairs at the same similarity of 0.15. Only for word pair 11 does Type2-Dist similarity measure produce a value close to 0.15.

All similarity measures also agree on at least one word pair with the greatest similarity value, word pair 9. This word pair is somewhat of a reasonableness check since the pair has identical words. But note that Type2-Dist also produced a similarity value of 1 for word pair 15. This result is due to the defuzzified mean value of the Type 2 interval fuzzy sets being basically identical for those two words *midpoint* and *centre* based on the human evaluations. For the ontology-based similarity measures, all three produced similarity values extremely close to 1 so that this word pair is also listed for them. Both FAST and FUSE have an additional word pair that produces a value of 1, word pairs 27 and 30, respectively. These results may be attributed to the difference in the construction of the ontology structures created using the defuzzified mean values for FAST and FUSE.

For the average similarity values, STASIS has the lowest one. This result is again expected since this similarity measure does not consider fuzzy words, only a word's position in the WordNet hierarchy. The Jaccard set-based measure follows closely after STASIS with the next lowest average. With the type-1 fuzzy set creation by human experts, the experts only provided one number in the [0, 10] interval and the standard deviations were based on the set of expert evaluations. It is possible that the triangular fuzzy sets created from the mean and standard deviation values are a poorer representation that affects the set-based fuzzy similarity measure more than the distance based GeoSim and partial matching Zadeh measures. More experiments are needed to verify this possible explanation for Jaccard's lower similarity values.

From Table I, comparison for producing highest similarity values among all similarity measures shows that FUSE produces the highest or ties for highest with FAST for 12 of the word pairs. FAST has the highest similarity or ties with FUSE 9 of the word pairs. Out of those word pairs with the highest similarity values, FAST and FUSE tie 6 times. FUSE produces higher similarity values because even for word pairs falling in the same node both within FAST and FUSE and, therefore, having a path length l equal to 0, the depth of the node h is typically at a higher level in FUSE than in FAST due to a maximum depth of 5 for FUSE compared to that of 2 for FAST. In addition the parameter β for FUSE is larger than that of FAST, i.e. 0.85 compared to 0.6. When FAST does produce a higher similarity, the path length l between the word pairs in FUSE's ontology is much greater than that in FAST's ontology, and with this case typically both word pairs are on different paths from the root node in both the FAST and FUSE ontologies. The depth h, therefore, would have the same value since the subsumer is the root node.

As can be seen from Table I, the fuzzy set based similarity measures rarely produce similarity measures greater than those that use the semantic similarity measure within an ontology. GeoSim and Type2-Dist have highest similarity for 2 word pairs each. Zadeh only has highest similarity once. For the lowest similarity values, Jaccard has lowest similarity for 12 of the 20 word pairs. It is to be expected that the semantic similarity measure used within the FAST and FUSE ontologies would produce higher similarity values than the fuzzy set similarity measures since there is a limit to the greatest path length of 4 and 10, and depth of 2 and 5, respectively The results from the semantic similarity measure are very much dependent on the structure of the ontologies that have been developed from the type-1 and type-2 interval fuzzy sets.

FUSE generally produces higher similarity values but both FAST and FUSE agree on numerous word pairs. This can occur when both the path distance l between the words pairs and the depth h of the subsumer of the word pairs are identical in both the FAST and FUSE ontologies.

V. CONCLUSIONS AND FUTURE WORK

This paper has conducted a study on fuzzy word sets derived from data collected from human participants and evaluates the performance of four simple fuzzy set similarity measures. It compares these results to the results of one semantic similarity measure as applied to two different ontologies created for FAST and FUSE from the fuzzy word sets. From the study, a very simple fuzzy set similarity measure created from COG distance between type-2 fuzzy sets has a very high correlation with the FUSE similarity results, even higher than that of FAST results with FUSE, both of which use the same semantic similarity measure. This result demonstrates that the construction of the ontology for the categories plays a significant factor in the resulting similarity values. The major difference between the two ontologies is in the level of detail considered in their construction. FAST is created using type-1 fuzzy sets and uses only 5 nodes with a depth of 2 in its ontology. FUSE is created using type-2 interval fuzzy sets and its ontology has 11 nodes with a depth of 5. Creating these ontologies is not straightforward and determining the appropriate structure for fuzzy word categories needs more investigation.

Although ontology creation for fuzzy words is challenging and it is unlikely that human perceptions of all the fuzzy words in a given language could be modelled, even with a limited number of fuzzy word models, the use of fuzzy semantic similarity measures in applications is beneficial. One aspect of future work looks at incorporating such measures into dialogue systems to replace traditional pattern matching algorithms with short text comparisons. Another area is to use the fuzzy set similarity measures instead of semantic similarity within the sentence similarity systems of FAST and FUSE to determine how well they correlate with human judgments. A hybrid of a fuzzy set similarity measure and a semantic similarity measure should be experimented with for the cases where sentence similarity does not agree with the human judgments of sentence similarity.

REFERENCES

[1] L. Zadeh, "From Computing with Numbers to Computing with Words—from Manipulation of Measurements to Manipulation of Perceptions. Logic, Thought and Action," International Journal of Applied Math. Comput. Sci., vol.12, no.3, pp. 307–324, 2002.

[2] G. Salton, C. Buckle, Term-weighting approaches in automatic textretrieval", Information processing & management vol.24, no. 5, pp.513-523, 1988.

[3] T. Landauer, P. Foltz, D. Laham, "An introduction to latent semantic analysis," Discourse processes vol. 25, no 3, pp.259-284,1998..

[4] Y. Li, D. Mclean, Z. Bandar, J. O'Shea, K. Crockett, "Sentence similarity based on semantic nets and corpus statistics", IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 8, pp.1138-1150, 2006.

[5] V. Cross, Xinran Yu, Xueheng Hu, "Unifying ontological similarity measures: A theoretical and empirical investigation," Int. J. Approx. Reasoning vol. 54 no. 7, pp. 861-875, 2013.

[6] Li, Y, Bandar, Z. McLean, D. "An approach for measuring semantic similarity between words using multiple information sources". IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 4, pp.871-882, 2003.

[7] D. Chandran, K. A. Crockett, D McLean, Z. Bandar, "FAST: A fuzzy semantic sentence similarity measure," International Conference on Fuzzy Systems, FUZZ-IEEE, 2013.

[8] N. Adel, K. A. Crockett, A. Crispin, D. Chandran, J. P. Carvalho, "FUSE (Fuzzy Similarity Measure) - A measure for determining fuzzy short text similarity using Interval Type-2 fuzzy sets," International Conference on Fuzzy Systems, FUZZ-IEEE pp. 1 -8 2018:

[9] Mendel, J. "Computing with words and its relationships with fuzzistics", Information Sciences vol. 177, no. 4, pp.988-1006, 2007.

[10] V. Cross, An Analysis of Fuzzy Set Aggregators and Compatibility Measures, Ph.D. Dissertation, Computer Science and Engineering, March 1993, Wright State University, Dayton, OH, 264 pages.

[11] P. Jaccard. "The distribution of the flora in the alpine zone", *New Phytologist*, vol. 11, pp. 37–50, 1912.

[12] V. Cross, T. Sudkamp, "Geometric compatibility modification," Fuzzy Sets and Systems, vol. 84, no. 3, pp. 283-299, 1996.

[13] M. Hao and J. M. Mendel, "Encoding words into normal interval type-2 fuzzy sets: HM approach," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 4, pp. 865-879, 2016.