



Please cite the Published Version

Cross, Valerie, Mokrenko, Valeria, Crockett, Keeley  and Adel, Naeemeh  (2020) Using Fuzzy Set Similarity in Sentence Similarity Measures. In: IEEE World Congress on Computational intelligence - IEEE FUZZ 2020, 19 July 2020 - 24 July 2020, Glasgow, UK (virtual congress).

DOI: <https://doi.org/10.1109/FUZZ48607.2020.9177836>

Publisher: IEEE

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/625507/>

Usage rights:  In Copyright

Additional Information: Copyright IEEE.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Using Fuzzy Set Similarity in Sentence Similarity Measures

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

Valeria Mokrenko
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
Department of Computing and
Maths, Manchester Metropolitan
University
Manchester, UK
N.Adel@mmu.ac.uk

Abstract— Sentence similarity measures the similarity between two blocks of text. A semantic similarity measure between individual pairs of words, each taken from the two blocks of text, has been used in STASIS. Word similarity is measured based on the distance between the words in the WordNet ontology. If the vague words, referred to as fuzzy words, are not found in WordNet, their semantic similarity cannot be used in the sentence similarity measure. FAST and FUSE transform these vague words into fuzzy set representations, type-1 and type-2 respectively, to create ontological structures where the same semantic similarity measure used in WordNet can then be used. This paper investigates eliminating the process of building an ontology with the fuzzy words and instead directly using fuzzy set similarity measures between the fuzzy words in the task of sentence similarity measurement. Their performance is evaluated based on their correlation with human judgments of sentence similarity. In addition, statistical tests showed there is not any significant difference in the sentence similarity values produced using fuzzy set similarity measures between fuzzy sets representing fuzzy words and using FAST semantic similarity within ontologies representing fuzzy words.

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

I. INTRODUCTION

Humans often find it easier to express domain knowledge using inexact, vague terms, or fuzzy words. Such words challenge the communication between humans and machines. Determining how similar two blocks of text are also faces the challenge of dealing with fuzzy words. Measuring the similarity between crisp words has typically been handled using semantic similarity measures. Much research has examined the use of semantic similarity measures within the context of an ontology, a knowledge structure containing concepts and defining the relationships between these concepts.

STASIS [1] is a system that produces sentence similarity measures between blocks of text. It measures the similarity between pairs of individual words, one from each block. The

semantic similarity measure proposed in [2] is used within the WordNet ontology. Although the STASIS work 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.

Fuzzy sets can serve as a means of representing fuzzy words. A framework for handling fuzzy words is the computing with words (CWW) [3] methodology by which fuzzy words can be quantified, scaled against each other and then become machine representable. The quantifying and scaling steps require that humans provide their perception of fuzzy words. Once fuzzy words are machine representable, then similarity measurement between the words can be performed. This fuzzy word similarity measurement is a necessary task in defining fuzzy sentence similarity measures (SSMs).

Since STASIS does not handle fuzzy words, additional research pursued improvements to SSMs by addressing this limitation. The FAST (Fuzzy Algorithm for Similarity Testing) [4] system uses CWW methods to develop a SSM that incorporates the similarity measurement between fuzzy words found in sentences or pieces of short text. The additional work in FAST to handle fuzzy words showed an improvement in its SSM as compared to that of STASIS when evaluated based on their correlations with human judgments of sentence similarities. This experimental study required the creation of datasets containing quantified fuzzy words. The quantification was based on surveying humans on their perceived numerical evaluation of the fuzzy words. These fuzzy words are organized into structured ontologies where the semantic similarity found in [2] can be used to measure the similarity between the fuzzy words.

Fuzzy sets can be modelled as type-1 or type-2 fuzzy sets. Further research has explored the use of type-2 fuzzy sets for fuzzy words in the FUSE (FUZZY Similarity mEasure) system [5]. It extends FAST by replacing the type-1 fuzzy set representations with type-2 fuzzy sets. The rationale was that type-1 fuzzy sets could not reflect the subjective nature of the

human evaluators and capture the uncertainty of humans [6]. Interval sets are used to represent the type-2 membership functions since they are simpler to use.

As done in FAST, the fuzzy words are arranged into ontologies. The same semantic similarity [2] used in STASIS and FAST is used in FUSE. Both FUSE and FAST require this step of transforming the fuzzy sets representing the fuzzy words into ontologies so that a semantic similarity measure can be used within the constructed ontologies. The major difference between the FAST and FUSE ontologies is in the level of detail considered in their construction. FAST with its type-1 fuzzy sets uses only 5 nodes with a depth of 2 in its ontologies. FUSE with its type-2 interval fuzzy sets uses ontologies with 11 nodes with a depth of 5. Building these ontologies based on the developed fuzzy sets for the fuzzy words is a required step to use the semantic or ontological similarity measure.

A previous paper [7] focused on determining if fuzzy set similarity measures might be used directly on the fuzzy sets with the goal of eliminating the ontology construction step. The measurement of similarity between fuzzy words represented as type-1 fuzzy sets used the following three existing fuzzy set similarity measures [8] of Zadeh's sup-min, Jaccard, and GeoSim. A fourth similarity measure referred to as Type-2 Dist uses a scaled COG for the type-2 fuzzy sets and the distance between their scaled COGs. The fuzzy set definitions for the fuzzy words for both type-1 and type-2 were obtained from the authors in [5]. The paper [7] reports on how well these simpler fuzzy set similarity measures correlated with the semantic similarity measure used in FAST and FUSE.

All of the fuzzy set similarity measures had a much higher correlation with FUSE's semantic similarity results based on its more sophisticated 11 node ontologies than FAST's semantic similarity results correlated with those of FUSE. The study showed that the results from the FAST and FUSE semantic similarity measures are very much dependent on the structure of the ontologies that have been developed from the type-1 fuzzy sets and type-2 interval fuzzy sets. Evaluating the use of these fuzzy set similarity measure in the computation of the FAST and FUSE sentence similarity measures, however, was not undertaken in that work.

The objective of this paper is investigate the performance of the fuzzy similarity measures when they replace the semantic similarity measure of FAST and FUSE in their SSMs. Their performance in the task of sentence similarity measurement is evaluated based on how well the resulting SSMs correlate with that of human judgments of sentence similarity.

The paper organization is as follows: Section II reviews from [8] the fuzzy word representation used and the four fuzzy set similarity measures. Section III describes the software that was available for this study and the modifications made to use the fuzzy set similarity measures in the SSMs. Section VI explains the experiments and how the evaluation of the fuzzy set similarity measures is performed. The results from the experiments and analysis are presented. Finally, Section V presents the conclusions and future work.

II. FUZZY WORDS AND SIMILARITY BETWEEN THEM

This section is a summary of the description found in [7]. To use fuzzy set similarity measures the fuzzy words must have a fuzzy set representation. The FAST research used questionnaires with human evaluators to develop a defuzzified value or mean and the standard deviation for each of the fuzzy words. These values for type-1 fuzzy sets were acquired from the FAST researchers. 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. Fig. 1 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 fuzzy 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.

The same twenty word pairs and the triangular membership type-1 fuzzy sets created for them in [7] are used for this research along with the associated pairs of sentences containing those fuzzy words. The first three fuzzy set similarity measures discussed below can simply be used on the triangular membership functions. Because for FUSE it was thought that type-2 fuzzy sets may better represent the subjective nature of a fuzzy word. Type-2 interval fuzzy sets were created for fuzzy words and a center of gravity (COG) was determined using the upper and lower footprints of uncertainty for the type 2 fuzzy sets. These COG values were acquired from the FUSE research and used in [7] as well as in this current research. The fourth fuzzy set similarity measure type-2 distance uses distance between COGs in determining the similarity between fuzzy words. Both the 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.

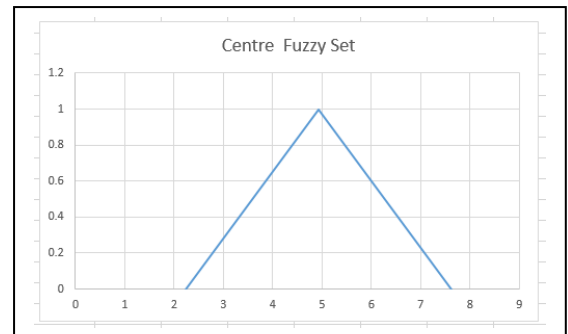


Fig. 1 Centre fuzzy set

A. Sup-Min

In [8] 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 matching index since it only provides an estimated similarity value between the two fuzzy sets.

B. Jaccard

The fuzzy Jaccard similarity measure is defined as a fuzzy extension of the Jaccard index 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 considered and not just the intersection points 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.

C. 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 [9], 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

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

where $[\beta_1, \beta_2]$ is an interval that contains both A' and A . 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 the distance between the two sets is used.

D. Similarity on Type-2 Defuzzified Values Distance

As previously explained in [5], type-2 interval fuzzy sets were used and then defuzzified into a single value by adapting Mendel's footprint of uncertainty (FOU) method [6]. 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:

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

The distance between the two centers of gravity is normalized by the size of the scaled interval $[-1, +1]$.

III. SOFTWARE USED AND MODIFIED IN EXPERIMENTS

In [7], the study focused on determining how well fuzzy set similarity measures correlated with the semantic similarity measure proposed in [2] which were used on the 5 node ontologies in FAST and the 11 node ontologies in FUSE. In that study, the effectiveness of the fuzzy set similarity measures in the overall task of determining sentence similarity was not investigated; only how well the different fuzzy words similarity measures correlated with each other is reported.

Here our research requires that the previous sentence similarity measurement systems, such as STASIS and FAST, be modified to use the fuzzy set similarity measures between fuzzy words in place of the semantic similarity measure. This modification is needed to determine the performance of fuzzy set similarity measures within a sentence similarity measurement. First STASIS, FAST, and FUSE are briefly described and then the modifications made first to STASIS and then to FAST are presented.

A. STASIS, FAST, and FUSE

STASIS [1] determines the degree of similarity between sentences or short blocks of texts by using both semantic information and word order information implied in the sentences. The semantic similarity calculation relies on the vocabulary in the WordNet ontology and on corpus statistics found in the Brown Corpus [10]. The semantic similarity between pairs of words w_1 and w_2 , one taken from each sentence, is determined as

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

where l represents the length of the path between the two words in WordNet and h represents the depth of their common subsumer.

STASIS uses this word similarity measure between all possible pairs of words from the two texts. A semantic vector is created for the sentence that weights the similarity based on the importance of the words where importance is derived from the Brown Corpus statistics. A syntactic vector is created using the positions of words in the texts. These two vectors are combined to produce an overall level of similarity for the two sentences.

FAST was developed to address the limitations of STASIS since STASIS is not able to determine semantic similarity between fuzzy words in judging the similarity between pairs of sentences. A major task to accomplish this is the creation of fuzzy set representations for the fuzzy words. A dataset containing quantified fuzzy words is organized hierarchically into six different categories [7]: *age*, *size/distance*, *frequency*, *goodness*, *membership level* and *temperature*. As previously explained, the fuzzy words in each category were quantified by human subjects. FAST used this quantification to create a 5 node ontology for each of the categories. FAST can be seen as an extension of STASIS since in order to use the same semantic similarity measure $S(w_1, w_2)$, FAST required these ontologies. FAST follows the same approach to measuring sentence similarity except fuzzy words that are not found in WordNet do not receive a similarity of 0. Instead the fuzzy words are found in the appropriate category ontology and their semantic similarity can be used in determining the overall sentence similarity.

FUSE is an extension of FAST in that it uses the same approach with organized category ontologies but they are built from type-2 fuzzy sets. These type-2 fuzzy sets were based on questionnaires completed by human subjects. FUSE also increases the fuzzy word vocabulary by 57%. The $S(w_1, w_2)$ measure is used in FUSE as done with STASIS and FAST. The only difference is the ontologies being used to determine path length l and the depth h of the common subsumer.

Both the STASIS and the FAST code are written in Python and were acquired from the first two authors in [5]. The provided FAST code was stated to be an earlier version of the code on which FUSE development was initiated. An effort was made to acquire more recent FAST code by contacting the first author of [4] in hopes of getting the most recent FAST. This effort was not successful. All results using this FAST code or modifications of FAST are from the use of this acquired FAST code; therefore, the SSM results reported in this paper could deviate from those reported in [4]. The FUSE code, also written in Python, was not made available, but the sentence similarities produced by FUSE were provided for the pairs of sentences used in this research. These sentence similarity values are given in the FUSE Reported column in Table VI.

B. Software Modifications

The objective was to modify STASIS so that instead of producing a fuzzy word similarity of 0 when the fuzzy words could not be found in WordNet, the fuzzy set similarity

measure between the fuzzy words is used. WordNet is used to check for synsets for a word in the sentence. If no synsets are returned, then the word cannot be found. The word may also have multiple synsets associated with it since a word could have more than one “sense” such as the word *bat*, i.e., the animal bat and a bat used in baseball. As implementation and testing progressed, it was discovered that some fuzzy words might be found in the WordNet ontology so that they could have a semantic similarity with another non-fuzzy word, i.e., a fuzzy and non-fuzzy word pair or may be paired with another fuzzy word. The implementation decision was made to only replace the fuzzy and non-fuzzy semantic similarity value with the fuzzy set similarity value for the fuzzy-fuzzy word pair if the fuzzy set similarity is greater.

A study of the FAST code revealed that the FAST developer made the same decision to replace the WordNet semantics similarity with the similarity produced by using the category ontologies for fuzzy words only if it was greater than the WordNet semantic similarity. This finding confirmed the decision that had been made in the modifications to STASIS to add fuzzy set similarity measurement between fuzzy words. The FAST code is very similar to the STASIS code except for the use of the category ontologies when determining similarity between fuzzy words. Here the modification required replacing the use of semantic similarity measures with fuzzy set similarity measures between the fuzzy words.

In both the obtained STASIS and FAST code, identical words are only assigned a similarity value of 1 if no synsets exist for the identical word. If a word has multiple synsets, as previously explained, it has multiple senses. If this is the case, then the assigned semantic similarity is the maximum between all the different senses for the word. As implementation progressed, a decision was made to create another version that also assigns a similarity value of 1 if there is only one synset for the word since that means there is only one “sense” for the word.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The FAST research developed pairs of sentences for evaluating the FAST sentence similarity measures. First a list of 30 sentence pairs from the dataset in [11] was generated with 20 having a high level of similarity, 5 of medium level and 5 of low. The sentence pairs were split and sentences randomly divided among three English language experts, who added a fuzzy word to each sentence and enhanced or reduced a particular attribute from it. From the three versions of the sentence, two were randomly paired together. Then the similarity of the sentences pairs were determined by surveying 18 people using questionnaires that asked them to rate how similar the sentences were on a scale of 0 to 10. These numbers were summarized as an average human rating score (AHR). Of the 30 sentence pairs, 20 contain fuzzy words in the same fuzzy category. These pairs with their italicized fuzzy words are listed in Table I and used in this study.

TABLE I. 20 SENTENCE PAIRS WITH FUZZY WORD

Sentence Pair	Sentence 1 Sentence 2
P1	When I was going out to meet my friends there was a <i>short</i> delay at the train station. The train operator announced to the passengers on the train that there would be a <i>massive</i> delay.
P5	Sometimes in a <i>large</i> crowd accidents may happen, which can cause life threatening injuries. There was a <i>small</i> heap of rubble left by the builders outside my house this morning.
P7	If you continuously use these products, I guarantee you will look very <i>young</i> . I assure you that, by using these products over a long period of time, you will appear almost <i>youthful</i> .
P8	I always like to have a <i>tiny</i> slice of lemon in my drink, especially if it's coke. I like to put a <i>large</i> wedge of lemon in my drinks, especially cola.
P9	I dislike the word quay, it confuses me every time, I <i>always</i> think of the thing for locks, there's another one. I dislike the word quay, it confuses me every time, I <i>always</i> think of the thing for locks, there's another one.
P10	Though it took many hours travel on the extremely <i>long</i> journey, we finally reached our house safely. We got home safely in the end, though it was a <i>mammoth</i> journey.
P11	The man presented a <i>minuscule</i> diamond to the woman and asked her to marry him. A man called Dave gave his fiancée an <i>enormous</i> diamond ring for their engagement.
P13	The <i>tiny</i> ghost appeared from nowhere and frightened the old man. The <i>diminutive</i> ghost of Queen Victoria appears to me every night, I don't know why, I don't even like the royals.
P15	Midday is 12 o'clock in the <i>midpoint</i> of the day. Midday is 12 o'clock in the <i>centre</i> of the day.
P16	The first thing I do in a morning is make myself a <i>lukewarm</i> cup of coffee. The first thing I do in the morning is have a cup of <i>hot</i> black coffee.
P18	This is a terrible noise level for a new car, I expected it to be of <i>good</i> quality. That's a very good car, on the other hand mine is <i>great</i> .
P19	Meet me on the <i>huge</i> hill behind the church in half an hour. Join me on the <i>small</i> hill at the back of the church in 30 minutes.
P20	It gives me <i>immense</i> pleasure to announce the winner of this year's beauty pageant. It's a <i>great</i> pleasure to tell you who has won our annual beauty parade
P22	Will I have to drive a <i>great</i> distance to get to the nearest petrol station? Is it a <i>long</i> way for me to drive to the next gas station?
P23	You have a very familiar face; do I know you from somewhere <i>nearby</i> ? You have a very familiar face; do I know you from somewhere where I used to live <i>faraway</i> .
P24	I have invited a <i>great</i> number of different people to my party so it should be interesting. A <i>small</i> number of invitations were given out to a variety of people inviting them down the pub.

P25	I am sorry but I can't go out as I have <i>loads</i> of work to do. I've a <i>gargantuan</i> heap of things to finish so I can't go out I'm afraid.
P27	Will you drink a glass of <i>excellent</i> wine while you eat? Would you like to drink this <i>wonderful</i> wine with your meal?
P29	<i>Large</i> boats come in all shapes but they all do the same thing. <i>Oversized</i> chairs can be comfy and not comfy, depending on the chair.
P30	I am so hungry I could eat a whole <i>big</i> horse plus desert. I could have eaten another <i>massive</i> meal, I'm still starving.

In Table II and Table III, several different sentence similarity results are given for STASIS and FAST. The second column in these two tables is the AHR for the pair of sentences. The Reported Results are those provided directly from the FAST and FUSE researchers. The Obtained Code Zero Synset column shows the results produced by running the code provided by the FAST and FUSE researchers. In this code identical words must have zero synsets to be assigned a similarity value of one. The last column shows the results for the modified STASIS and FAST code that uses a test checking if identically spelled words have only one or zero synsets.

Table II and Table III also show both the Pearson and Spearman correlation of the various STASIS and FAST versions. A substantial difference in the correlation with the similarities of these versions to the AHR exists. For STASIS, the Obtained Code results highest correlation. A possible explanation for the difference in the Reported versus the Run results is the difference in the Natural Language Took Kit (NLTK) [12] versions from when STASIS originally produced the results and the current version (3.4.5). It is unlikely WordNet versions caused the difference because although synset offsets change, synsets are stable with few splits and merges between the versions [13]. Similarly for FAST, the Obtained Code results produce a higher correlation than the Reported results. ANOVA analysis performed on the three different versions of STASIS indicates the means for three STASIS versions are not significantly different at the 0.05 level. The same outcome occurred for the three FAST versions.

FAST correlations are higher than the STASIS correlations. These higher correlations are expected since FAST handles fuzzy words that STASIS is not able to. To determine if a statistically significant difference exists in the means of the STASIS and FAST Obtained Code results, a two tail t-test was performed. The outcome of the t-test shows that there is a significant difference between STASIS Obtained Code Results and FAST Obtained Code Results at the 0.05 level with p-value of 0.0073. For the t-test between STASIS Modified Zero or One Synset and FAST Modified Zero or One Synset produced a significant difference with a p-value of 0.0013. These t-tests verify that handling fuzzy words in FAST does significantly improve correlation with human judgments of sentence similarity over that of STASIS.

TABLE II. STASIS RESULTS

Sentence Pairs	Human Judgment (AHR)	Reported Results	Obtained Code Zero Synset	Modified to Zero or One Synset
P1	3.833	0.74688	0.74688	s 0.746352
P5	1.281	0.553945	0.553945	0.543552
P7	7.095	0.854431	0.854431	0.806182
P8	6.719	0.90160	0.779976	0.763487
P9	0.952	0.68323	0.615611	0.619075
P10	8.248	0.707534	0.707534	0.742276
P11	4.957	0.531135	0.465735	0.449818
P13	3.286	0.533853	0.564408	0.577580
P15	9.138	0.999921	0.999926	0.999889
P16	6.781	0.844044	0.844044	0.844044
P18	2.11	0.475089	0.496757	0.496756
P19	6.757	0.779292	0.779292	0.732798
P20	8.986	0.758728	0.823382	0.793999
P22	8.852	0.882129	0.882129	0.881933
P23	7.043	0.858609	0.858609	0.858609
P24	3.833	0.707128	0.707128	0.707051
P25	8.857	0.626350	0.742006	0.693284
P27	8.919	0.707795	0.707795	0.614119
P29	1.295	0.389489	0.389489	0.268906
P30	6.624	0.508935	0.534416	0.529974
Pearson with AHR		0.631977	0.724682	0.6775326
Spearman with AHR		0.628056	0.717563	0.642347

The previous discussion compared the results of STASIS and FAST without using fuzzy set similarity measures. FAST handles fuzzy words by using semantic similarity within its category ontologies. Table IV show the sentence similarity values produced after modifying the obtained STASIS code to use each of the four fuzzy set similarity measures between fuzzy words that cannot be found in WordNet. The fuzzy set similarity value may also replace the semantic similarity measure from WordNet when the fuzzy set similarity value is greater than that of the semantic similarity within WordNet. The values in Table IV are based on using a check for Zero or One Synset when identical words are found. That version produces slightly higher correlations with human judgments of sentence similarities than just checking for zero synsets.

Since STASIS GeoSim has the highest Pearson correlation with human judgments, a t-test on the SSM values between it and those of FAST Modified Zero or One Synset was performed. The purpose of the t-test is to determine if the modified STASIS using fuzzy set similarity measures differs significantly from FAST using semantic similarity within its category ontologies. There is no statistically significant difference between their SSM values with a p-value of 0.06

TABLE III. FAST RESULTS

Sentence Pairs	Human Judgment (AHR)	Reported Results	Obtained Code Results	Modified to Zero or One Synset
P1	3.833	0.716059	0.766476	0.76603505
P5	1.281	0.553945	0.554011	0.54362499
P7	7.095	0.848375	0.837837	0.83783737
P8	6.719	0.896886	0.772687	0.77268639
P9	0.952	0.681290	0.613872	0.61774001
P10	8.248	0.824531	0.822187	0.83205864
P11	4.957	0.517416	0.489134	0.4891348
P13	3.286	0.583988	0.608148	0.60813224
P15	9.138	0.999921	0.999890	0.99989027
P16	6.781	0.897493	0.861690	0.86169039
P18	2.11	0.498348	0.498887	0.49888599
P19	6.757	0.782346	0.779151	0.77914815
P20	8.986	0.782177	0.831654	0.83164902
P22	8.852	0.901850	0.900176	0.89987811
P23	7.043	0.891414	0.872690	0.87269049
P24	3.833	0.712779	0.713812	0.71373476
P25	8.857	0.664910	0.758325	0.75585767
P27	8.919	0.794916	0.792803	0.7927075
P29	1.295	0.372960	0.477078	0.37296983
P30	6.624	0.563401	0.567695	0.56340074
Pearson with AHR		0.718752	0.782514	0.785599
Spearman with AHR		0.686724	0.786762	0.780745

Table V shows the SSM values produced after modifying the obtained FAST code to use the four fuzzy set similarity measures between fuzzy words instead of the FAST semantic similarity measure within its category ontologies. Again, the SSM values in Table V are based on using Zero or One Synset check when identical words are found. Since the FAST Zadeh SSM values in Table V have the highest correlations with those of human judgments, a t-test between the SSM values of FAST Zadeh and those of FAST Modified Zero or One Synset found in Table III was performed. There is no statistically significant difference between the FAST Modified Zero or One Synset SSM values and the FAST Zadeh SSM values with a p-value of 0.30.

An ANOVA test was performed on the STASIS SSM values for the four fuzzy set similarity measures in Table IV and showed no statistically significant difference among the four fuzzy set similarity measures with a p-value of 0.98.

An ANOVA test was also performed on the FAST SSM values for the four fuzzy set similarity measures in Table V and showed no statistically significant difference among the four fuzzy set similarity measures with a p-value of 0.91.

TABLE IV. STASIS RESULTS WITH FUZZY SET SIMILARITY MEASURES

Sentence Pairs	GeoSim	Zadeh	Jaccard	Type2-Dist
P1	0.787064	0.763680	0.754448	0.775012
P5	0.544365	0.543552	0.543552	0.543552
P7	0.846477	0.845563	0.82348 S7	0.850371
P8	0.792185	0.779492	0.768859	0.787140
P9	0.623408	0.623408	0.623408	0.623408
P10	0.834697	0.833500	0.810405	0.836665
P11	0.498857	0.449818	0.449818	0.468732
P13	0.602700	0.611673	0.557758	0.612318
P15	0.999889	0.999899	0.999889	0.999903
P16	0.898757	0.903868	0.878825	0.886171
P18	0.512070	0.511553	0.504335	0.514108
P19	0.790848	0.770028	0.742340	0.760205
P20	0.834148	0.834457	0.828792	0.834934
P22	0.901838	0.901839	0.899851	0.901530
P23	0.914425	0.873909	0.866268	0.912640
P24	0.717015	0.711358	0.707051	0.710872
P25	0.795177	0.799607	0.758452	0.777154
P27	0.798399	0.803035	0.765686	0.798276
P29	0.443928	0.452315	0.413302	0.435904
P30	0.559292	0.559306	0.559001	0.558381
Pearson with AHR	0.780550	0.7769880	0.766168	0.771685
Spearman with AHR	0.807823	0.805566	0.803310	0.792779

TABLE V. FAST RESULTS WITH FUZZY SET SIMILARITY MEASURES

Sentence Pairs	GeoSim	Zadeh	Jaccard	Type2-Dist
P1	0.77846842	0.73734862	0.71299107	0.75310801
P5	0.53784191	0.53702486	0.53702486	0.53702486
P7	0.77651531	0.77548316	0.75288337	0.78118585
P8	0.79134429	0.77900059	0.76880401	0.78646828
P9	0.6249823	0.6249823	0.62498230	0.62498230
P10	0.78147147	0.77990048	0.75305702	0.78415594
P11	0.49646667	0.48913428	0.48913428	0.46645995
P13	0.59921112	0.60815523	0.55496662	0.60880177
P15	0.99989028	0.99990028	0.99989028	0.99990456
P16	0.89211791	0.86003185	0.76476191	0.91034563
P18	0.53525255	0.53306474	0.50409044	0.54414839
P19	0.79157590	0.77103232	0.74372152	0.76134248
P20	0.83366025	0.83397266	0.82830291	0.83445445
P22	0.83924665	0.83910880	0.83429783	0.84016202
P23	0.85950594	0.79375393	0.77808485	0.85468583
P24	0.69718287	0.69171208	0.68753593	0.69124231
P25	0.75802306	0.76293300	0.72051962	0.73925663
P27	0.79671784	0.80137271	0.76401735	0.79659489
P29	0.44978550	0.45828609	0.41897634	0.44168157
P30	0.56064827	0.56055748	0.54297103	0.56498333
Pearson with AHR	0.76525447	0.78809245	0.77830218	0.75649209
Spearman with AHR	0.7589320	0.823618	0.8115834	0.782249

Table VI shows the Reported Results for FUSE on the 20 pairs of sentences. The FUSE code could not be obtained so only the reported results are provided. FUSE's Pearson correlation with the human judgments of SSM values is greater than 0.631977 of the STASIS Reported Results. This result is to be expected since STASIS does not handle fuzzy words. It is, however, slightly lower than the 0.718752 of the FAST Reported results. A t-test was performed between Reported FUSE SSM values and Reported FAST SSM values to see if this difference is statistically significant. The t-test result indicates their SSM values for the two are not statistically significant with a p-value of 0.74.

The SSM values of the Fuse Reported Results are compared to those of STASIS GeoSim using a two tailed t-test. The result showed no statistically significant difference in their SSM values with a p-value of 0.55. The SSM values of the Fuse Reported Results are also compared to those of FAST Zadeh using a two tailed t-test. The result showed no statistically significant difference in their SSM values with a p-value of 0.21.

TABLE VI. FUSE RESULTS

Sentence Pairs	FUSE REPORTED
P1	0.736759
P5	0.553945
P7	0.802018
P8	0.896688
P9	0.674952
P10	0.781539
P11	0.530601
P13	0.590667
P15	0.999921
P16	0.892987
P18	0.480010
P19	0.753917
P20	0.784651
P22	0.886215
P23	0.908339
P24	0.707962
P25	0.656508
P27	0.792183
P29	0.538658
P30	0.582188
Pearson with AHR	0.691786
Spearman with AHR	0.68317

V. CONCLUSIONS AND FUTURE WORK

FAST [4] developed a method to handle fuzzy words in the measurement of sentence similarity, and FUSE [5] later made enhancements for handling fuzzy words. Both of these approaches arrange fuzzy words into category ontologies and use semantic similarity measures within the ontologies to determine the similarity between fuzzy words. This paper presents a study on the use of fuzzy set similarity measures in place of semantic similarity measures within ontologies.

The results of the experiments with the modified STASIS and FAST code show that the modifications to these two to use fuzzy set similarity measures produce SSMs with correlations very close to and just as good as those correlations that STASIS and FAST produced using their semantic similarity measures. In particular, Zadeh's fuzzy set similarity measure, when used in the modified FAST code, produced both the greatest Pearson and Spearman correlations as see in Table V. Using a t-test between the two approaches for the SSM values that produced the highest correlations for each approach showed there is no statistically significant

differences in the SSM values between the FAST obtained code and the FAST code modified to use fuzzy set similarity measures.

The main advantage of using fuzzy set similarity measures is that building the category ontologies in order to use the semantic similarity measure is not necessary. Fuzzy set similarity measures can be used directly on the fuzzy set representations of the fuzzy words. In the future, modifications to the FUSE code may also determine if using fuzzy set similarity measures could improve correlation with human judgment and may include an investigation into other type-2 fuzzy set similarity measures. In addition, a study of the different NLTK versions used to perform the SSM calculation may also be undertaken.

REFERENCES

- [1] 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.
- [2] Y. Li, Z. Bandar, D. McLean, "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.
- [3] 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.
- [4] D. Chandran, K. A. Crockett, D. McLean, Z. Bandar, "FAST: A fuzzy semantic sentence similarity measure," International Conference on Fuzzy Systems, FUZZ-IEEE, 2013.
- [5] 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:
- [6] Mendel, J. "Computing with words and its relationships with fuzzistics", Information Sciences vol. 177, no. 4, pp.988-1006, 2007.
- [7] V. Cross, V. Mokrenko, K. Crockett, N. Adel, "Ontological and Fuzzy Set Similarity between Perception-Based Words," International Conference on Fuzzy Systems, FUZZ-IEEE, 2019.
- [8] 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.
- [9] V. Cross, T. Sudkamp, "Geometric compatibility modification," Fuzzy Sets and Systems, vol. 84, no. 3, pp. 283-299, 1996.
- [10] Brown Corpus Information, http://www.essex.ac.uk/w3c/corpus_ling/content/corpora/list/private/brown/brown.html
- [11] O'Shea, J, Bandar, Z. Crockett, K, Mclean, D., "Pilot Short Text Semantic Similarity Benchmark Data Set: Full Listing and Description" Technical Report Available: www2.docm.mmu.ac.uk/STAFF/J.Oshea/TRMMUCCA20081_5.pdf
- [12] <https://www.nltk.org>
- [13] E. Kafe, "How Stable are WordNet Synsets?." *LDK Workshops*. 2017.