DOI: https://doi.org/10.1109/FUZZ-IEEE.2019.8858880

Publisher: IEEE

Version: Accepted Version

Human Hedge Perception – and its Application in Fuzzy Semantic Similarity Measures

Naeemeh Adel, Keeley Crockett, Alan Crispin
School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Chester Street, Manchester, M1 5GD, UK
N.Adel@mmu.ac.uk

Joao P. Carvalho
INESC-ID / Instituto Superior Tecnico, Universidade de Lisboa, Portugal

David Chandran
Institute of Psychiatry, Psychology & Neuroscience, Kings College London, 16 De Crespigny Park, London, SE5 8AF, UK

Abstract - **Fuzzy Semantic Similarity Measures are algorithms that are able to compare two or more short texts that contain human perception based words and return a numeric measure of similarity of meaning between them. Such similarity is computed using a weighting, comprised of the semantic and the syntactic composition of the short text. Similarities of individual words are computed through the use of a corpus, and ontological structures based on both WordNet – a well-known lexical database of English, and on category specific fuzzy ontologies created from the derivation of Type-I or Type-II interval fuzzy sets from human perceptions of fuzzy words. Currently, linguistic hedges are not utilized in the similarity calculation within fuzzy semantic similarity measures and are ignored. This paper describes a study, which aims to capture human perceptions for linguistic hedges typically used in natural language. Twelve linguistic hedges used within natural language are selected and an experiment is conducted to capture human perceptions of the impact of hedges on fuzzy category words. A dataset of hedge sentence pairs is created and rated in terms of similarity by human participants. Excellent inter-rater correlations and inter-class correlations are established between the average human ratings and an established fuzzy semantic similarity measure.**

Keywords— hedges, lingustic variables, fuzzy semantic similarity measures, interval type-II

I. Introduction

In the field of fuzzy logic, linguistic variables are a well-defined concept where the value of the variables are words that are used in natural human language [1]. In [1], Zadeh defines a term-set for each linguistic variable (i.e. Age) which constitutes all its possible numerical values i.e. [0..130], with linguistic values (i.e. *young*) acting as labels for fuzzy restrictions based upon the meaning and interpretation by a human in a given context. A number of linguistic hedges, designed to modify fuzzy sets through concentration, intensification and dilation, were first defined as mathematical models. In some cases there is not an agreement on the model [2, 3, 4] and more recent work by Le and Tran [5] on dual hedges (i.e. hedges can be used simultaneously to express different levels of emphasis) consider extensions to fuzzy logics through two axiomatizations for multiple hedges [5]. Novak [4] proposes that the branch of fuzzy logic, known as Fuzzy Natural Logic provides a rational model of linguistic semantics, and argues that hedging is more complex than previously known when applied within the field of linguistics. Current work on hedges includes the use of linguistic terms with weakened hedges (LTWH) to enhance natural uncertainty in decision-making [6, 7]. This work uses two frequently used hedges within qualitative decision-making, and argues that the formulation of more complex linguistic expressions improves decision making under uncertainty. Work in [7, 8] acknowledges that more linguistic hedges need to be determined especially for use within modelling natural language. The effect of hedges applied to fuzzy systems has been studied in many application domains such as enhancing a student's academic evaluation [8], the selection of a supplier based on a number of live parameters within a product's supply chain in small and medium businesses [9] and vehicular traffic density estimation [10]. However, their effect within the application of fuzzy semantic similarity measures has not been studied.

Fuzzy Natural Language Processing (FNLP) can be addressed with the formulation of fuzzy computational models of words [4]. We define fuzzy words, within this work, as any word that has a subjective meaning in natural language and is based on a human's perception in a given context. Fuzzy words are often defined from the bottom up – based upon obtaining a representative sample of the human population for a given word and context and then modelling the range of perceptions using either a Type-I or Type-II fuzzy set representation [11, 12]. This process of using humans' subjective opinions is adopted from the field of natural language processing [13] and from work undertaken by Mendel [14, 15, 16], first in his code-book using a Type-I representation and then following the Hao-Mendel Approach (HMA) using Interval Type-2 fuzzy sets [17].

The motivation for the work in this paper stems from a weakness in the application of fuzzy semantic similarity measures (FSSM) which are used to find a measure of the semantic and syntactic similarity, between short texts, typically of 25 words or less [18]. Currently, linguistic hedges are not utilised in the similarity calculation within FSSMs. Two such FSSM measures are FAST [11] and FUSE [12]. FAST was the first FSSM built on a limited number of categories of words represented by Type-I fuzzy sets used to derive category ontologies similar to WordNet [19]. FUSE (FUzzy Similarity mEasure) determined similarity using expanded categories of perception based words that were modelled using Interval Type-2 fuzzy sets [12]. We hypothesise that the inclusion of the semantic meaning of linguistic hedges will improve the precision of the similarity measurement through obtaining a higher correlation of similarity with human ratings. Hence, linguistic hedges are expected to make a weighted contribution when calculating the overall semantic similarity.

This paper is organized as follows; Section II provides a brief summary of background work on hedges and related work on FSSMs. Section III defines the study that aims to capture human perceptions for linguistic hedges typically used in natural language. In section III, the methodology for natural language hedge selection and obtaining human perceptions of hedges in relation to fuzzy words is described. Following the modelling of the hedges using Type-II interval fuzzy sets, the methodology for creating 16 hedge sentence pairs is presented. Section IV presents the results obtained from capturing perceptions of humans for 12 hedge words and obtaining human similarity ratings between hedge sentence pairs. Section V explores further work in exploring hedge weightings within FSSMs.

II. BACKGROUND AND RELATED WORK

A) Hedges

A linguistic variable carries with it the concept of fuzzy set qualifiers, called hedges. A hedge is a marker of uncertainty in language. Hedges are terms that modify the shape of fuzzy sets. They include adverbs such as *very*, *somewhat*, *quite*, *more or less* and *slightly* [20]. Linguistic variables represent crisp information in a form, and precision, appropriate for the problem. Linguistic variables associate a linguistic condition with a crisp variable. A crisp variable is the kind of variable that is used in most computer programs: an absolute value. A linguistic variable, on the other hand, has a proportional nature: in all of the software implementations of linguistic variables, they are represented by fractional values in the range of 0 to 1 [21]. Hedges can modify verbs, adjectives, adverbs or even whole sentences. They are used as [20]:

- All-purpose modifiers, such as *very*, *quite* or *extremely*
- Truth-values, such as *quite true* or *mostly false*
- Probabilities, such as *likely* or *not very likely*
- Quantifiers, such as *most*, *several* or *few*
- Possibilities, such as *almost impossible* or *quite possible*.

Hedges act as operations themselves. For instance, *very* performs concentration and creates a new subset from the fuzzy set it is applied to i.e. applying the hedge *very* to the set of *tall men*, derives the subset of *very tall men*. Hedges are useful as operations, but they can also break down continuums into fuzzy intervals. For example, the following hedges could be used to describe temperature: *very cold, moderately cold, slightly cold, neutral, slightly hot, moderately hot* and *very hot*. Obviously, these fuzzy sets overlap. Hedges help to reflect human thinking, since people usually cannot distinguish between *slightly hot* and *moderately hot* [20]. This makes them important when measuring human perceptions of the similarity of short texts.

According to Zadeh [22], a linguistic variable is a variable, whose values are words or sentences in a natural or artificial language, as opposed to numerical values. Therefore for the category *Age,* it would be considered a linguistic variable if its values were linguistic rather than numerical, this means A*ge* = {*young, not so young, very young… old, not very old, not very young*} is a linguistic variable, as opposed to *Age* = {20, 21, 22 … 60, 61…} which is a numerical variable.

A linguistic variable is characterised by a quintuple (*L*, *T*(*L*), *U, G, M*) where [22]:

- *L* is the name of the linguistic variable
- *T*(*L*) is the term set of *L* (collection of linguistic values)
- *U* is the universe of discourse
- *G* is a *syntactic rule* which generates the terms in *T*(*L*)
- *M* is a *semantic rule* which associates with each linguistic value *X* its meaning *M*(*X*)
- Where *M*(*X*) denotes a fuzzy subset of *U*.

Considering the example of *tall men*, application of the concentration hedge, *very* operation, will reduce the degree of memebership of fuzzy elements [20]. The application of hedge *very*, can be calculated using a mathematical square as follows:

$$\mu_A^{very}(x) = [\mu_A(x)]^2 \tag{1}$$

Thus if a person had a 0.84 membership in the set of *tall men*, then they will have a 0.7056 membership in the set of *very tall men*.

B) Fuzzy Semantic Similiarty Measures

Traditionally, Semantic Similarity Measures stemmed from the field of natural language processing and are used for measuring the degree to which a sentence or short-texts are subjectively evaluated by humans to assess whether or not they are semantically similar to each other. Traditional measures did not capture the use of fuzzy words - words that have subjective meanings to different people in different contexts, are typically ambiguous and are characteristically used in everyday human natural language dialogue [12]. The FAST algorithm (Fuzzy Algorithm for Similarity Testing) [1], is an ontology based similarity measure that uses concepts of fuzzy words represented by Type-I fuzzy sets. However, Type-I fuzzy sets were not able to correctly model the subjective options of humans on the meanings of fuzzy words in different contexts. FUSE, attempted to overcome this problem, by using Interval Type-II fuzzy sets to model relationships between categories of human perception based words using fuzzy category ontologies. The FUSE algorithm which can be found in [12], consisted of both syntactic and semantic components which were weighted. FUSE was able to model intra-personal (the uncertainty a person has about the word) and inter-personal (the uncertainty that a group of people have about the word) uncertainties, which are intrinsic to natural language. In [11], FUSE gave better correlations compared to human ratings than FAST over three benchmark datasets [16]. In these results, the modelling of linguistic hedges and the impact on the similarity measurement value was not considered. Hedges were not represented in the fuzzy category ontologies and therefore did not form part of the similarity measurement.

III. Capturing Human Perceptions of Hedges – A Study

A) Overview of study

The aim of this study is to investigate the effect of inclusion of hedge modifiers within the similarity calculation of fuzzy sentence similarity measures. The hypothesis is that their inclusion will improve the precision of the similarity measurement through obtaining a higher collaboration of similarity with human ratings. To investigate the hypothesis, a study consisting of two experiments was undertaken. The first experiment was to obtain human perceptions of the intensity that a hedge had on a fuzzy word. Fuzzy intensity in this research refers to the perceptive numerical measure a word is given, be that measure positive, or negative by a human rater.

For this experiment, let the fuzzy subset *Hedges = {Below, Approximately, Neighbouring, Roughly, About, Around, Quite, Indeed, Definitely, Positively, Very, Above}.* The fuzzy words were selected from the 6 original categories proposed in FUSE [12] as follows: *{Adequate (Level of Membership), Satisfactory (Worth), Middle-Aged (Age), Mild (Temperature), Fair (Frequency), Average (Size & Distance)}.*These fuzzy words were chosen by selecting the word with the value closest to 0 in each category on a scale of [-1, +1]. Once human perceptions were captured they could be used to construct Type-II interval models similar to those used in FUSE [12] and used to derive a hedge ontology. The ontology would be used to determine the path length and depth between words as part of the word component similarity measures in FUSE. The path length and depth of hedge words are relative to their position in the hedge ontology where each hedge category is treated as a concept. Each concept is constructed using a taxonomy (binary tree) where the root node always takes the value 0. Defuzzified hedge words are then placed into tree nodes at intervals of ± 0.2 [12] From the hedge taxonomy, the path length and depth of the Lowest Common Subsumer can be determined for hedge words in a category. This would allow the defuzzified hedge value to influence its associated defuzzified fuzzy word values, in terms of intensity, be this positively in that the sentence similarity value increased or negatively in that the sentence similarity value decreased.

B) Hedge Intensity Experiment

To determine intensity of hedges when applied to fuzzy words, 32 participants consented to take part in a study, all of whom were native English speakers above the age of 18. In total there were 12 hedge words that were not already present in the FUSE Fuzzy Dictionary [12] that had mathematical definitions. When the mathematical value of a hedge word, (such as *Very* as defined in *Eq. (1)*) was applied to a fuzzy word it did not represent the mathematical model that was linguistically represented, therefore a different approach was needed to cater for hedge words. As an example, the hedge word *Very* has a mathematical equation of x^2 [23], where x is the fuzzy value. Therefore taking the word *Hot=0.6193*, and computing the phrase *Very Hot*= $(0.6193)^2 = 0.3836$, calculated the mathematical value of *Very Hot* to be smaller than the mathematical value of Hot, whereas linguistically *Very Hot* has a more positive intensity then *Hot*. Therefore a different approach to measuring the intensity was required that required the perceptions of humans. To achieve this the subset of 12 hedge words where each added prior to the fuzzy words, one from each of the 6 categories represented in the FUSE FSSM [12]. The middle word in each category with the value closest to zero was selected, and a random hedge word was added to the beginning of each of these six words. Participants were first given a description of the task, which included a simple linguistic definition of a hedge and a fuzzy word. An extract from the experiment description is as follows: *"The aim of this experiment is to help contribute towards computer systems that will understand the English language. This experiment is about HEDGES. Hedges are terms that modify the shape of a sentence. They include adverbs such as very, somewhat, quite, more or less and slightly. In this experiment, I am going to give you 6 words belonging to 6 categories. A category in this instance is just the name given for a group of words that fall under a similar meaning. For instance, for the category TEMPERATURE, it will contain words such as [hot, cold, mild, boiling, scorching, freezing…]. I am going to give you a scale of 0 to 10. Each word sits in the middle of this scale (5). I am going to pair each word with some hedge words and would like you to tell me where these new words would sit on this scale. You can use one decimal place (e.g. 3.2) for finer precision."*

An image of a ruler (Figure 1) was used as a visual aid to make understanding the word placement visually easier. The chosen word from each category was always located at mark 5 on the ruler and was highlighted in red. The participants were then asked to rate the new *hedge word* when applied to the fuzzy *word* on this ruler on a scale of [0-10] with 1 decimal place permitted for accuracy. One example of a word used in this experiment is the hedge word *Below*. Taking the fuzzy word *Fair*, belonging to the category *Frequency*, one participant felt that the word *Below Fair* would be represented by a value of 3.4 as shown in Figure 1. Their opinion was that the hedge, *Below*, negatively reduced the intensity of the category word Fair.

The aim of the hedge intensity experiment was to try and mimic the perceptions of humans using natural language, despite them not actually thinking about words on a scale. On obtaining all human measurements, the average value for each hedge word was calculated and this was scaled on a scale of [-1, +1] to create a hedge ontology. This was done to match the same scale and ontological structure as the words in the fuzzy dictionary used within FUSE [12].

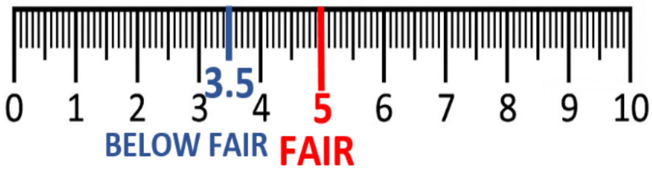

Fig. 1 - Scale for Hedge Intensity Experiment

C) *Human Ratings of Hedged Sentence Pairs*

In order to assess the intensity of hedges in the natural language context, it was necessary to compute the sentence similarity between pairs of sentences, which contained hedge words. Following analysis, it was established that the fuzzy sentence benchmark datasets, known as SFWD and MFWD [12], did not contain a sufficient number of hedge words in order to conduct a rigorous evaluation. Therefore, a dataset containing 16 sentence pairs containing hedge words was created. The methodology comprised of randomly extracting 16 sentences pairs from the MFWD [12] ranging from high to low similarity based on human ratings [12]. For each fuzzy word in the hedge sentence pair (HSP), a hedge word was assigned prior to that fuzzy word, i.e. for *HSP1* "The little village of Resina is also situated *approximately* near the spot", the hedge *approximately* was added. The sentence pairs were then checked by an English language expert, to ensure they were grammatically correct. Table I shows the full set of hedge sentence pairs.

O'Shea et. al. [13] emphasized the importance of establishing rigorous methodology when obtaining human ratings of similarities between words and sentence pairs, especially in relation to sample size, population distribution and the inclusion of calibration pairs providing representation of the highest and lowest sentence similarity pairs within the data set. Adopting this methodology, the second experiment consisted of 16 participants who were all native English speakers above the age of 18 from a diverse range of backgrounds. They were provided with the 16 HSPs and were asked to rate each sentence on a scale of [0-10], with 1 decimal place permitted for accuracy, based on how similar they were to each other. The scale of [0-10] was adopted to be consistent with approaches in [11, 12, 13, 24].

TABLE I HEDGE SENTENCE PAIRS CONTAINING FUZZY & HEDGE WORDS

Hedge Sentence Pairs	Sentence 1	Sentence 2
HSP 1	The little village of Resina is also situated approximately near the spot	He seems quite excellent man and I think him uncommonly pleasing
HSP 2	A little quickness of voice there is which definitely rather hurts the ear	The only living thing near was a very old bony grey donkey
HSP 3	It is as long again as approximately almost all we have had before	was scarcely less below warm than hers and whose mind -- Oh
HSP 4	A positively frosty youthful man	A indeed hot old man
HSP 5	A definitely thick juvenile man	A very little old man
HSP 6	Had you married you must have been quite regularly acceptable	Had you married you must have been indeed always poor
HSP 7	So would roughly useless diminutive Harriet	So would indeed poor little Harriet
HSP 8	Have massive mercy on the above mediocre men	Have a little mercy on the below poor men
HSP 9	How positively marvellous middling Piccola must have been	How quite good poor Piccola must have been
HSP 10	Behold how definitely fine a matter an adjacent fire kindleth	Behold how approximately great a matter a little fire kindleth
HSP 11	We will not say how small for fear of shocking the very youthful ladies	We will not say how indeed near for fear of shocking the young ladies
HSP 12	What s the fine roughly pensionable man	What's the roughly good old man
HSP 13	And he laughed around almost dreadfully	And he laughed very rather unpleasantly
HSP 14	Yesterday's ruling is a positively great first step toward better coverage for poor Maine residents he said but there is more to be done	He said the court 's ruling was a positively great first step toward better coverage for poor Maine residents but that there was more to be done.
HSP 15	It is largely a quite sizeable story, said Turnbull smiling	It is roughly rather a long story, said Turnbull smiling
HSP 16	The eyes were full of a frosty and quite frozen wrath a kind of utterly heartless hatred	The eyes were full of a frozen and quite icy wrath a kind of utterly heartless hatred

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A) *Hedge Intensity Results*

Table II shows the results of the Average Human Ratings (AHR) for the hedge intensities. The table shows the 6 words from the fuzzy dictionary categories (Fuzzy Words), and the 12

TABLE II. AHR FOR HEDGE INTENSITIES ORDERED HIGH TO LOW

Hedge Words \ Fuzzy Words	Adequate	Satisfactory	Middle-Aged	Mild	Fair	Average	Total Average	Scaled
Below	3.2500	3.5167	3.7176	3.7750	3.6118	3.5765	3.4885	-0.3023
Approximately	4.3273	4.4700	5.0083	4.8688	5.1529	4.8125	4.7813	-0.0437
Neighbouring	4.6615	5.0357	4.7231	4.7357	4.6923	4.8308	4.8031	-0.0394
Roughly	4.8192	4.8286	4.6273	4.4636	4.9692	4.3143	4.8036	-0.0393
About	5.0333	4.8643	5.1235	4.6231	5.0545	4.8556	4.8865	-0.0227
Around	4.8400	4.7632	4.9895	4.7889	4.6235	4.8238	4.9000	-0.0200
Quite	5.5353	5.5889	5.5500	4.6071	5.5458	5.6000	5.2943	0.0589
Indeed	5.8133	5.6600	5.9333	4.8867	6.2200	5.1600	5.3125	0.0625
Definitely	5.9000	6.9333	5.9000	5.6526	5.5150	5.6818	5.4573	0.0915
Positively	5.6154	6.4600	6.5333	6.2000	6.1313	5.9067	5.7823	0.1565
Very	6.8563	7.0533	6.9133	5.1563	6.8063	6.6250	6.4250	0.2850
Above	6.5353	6.6250	6.4375	6.2188	6.4375	6.6875	6.4854	0.2971

TABLE IIIA. COMPARISON OF HEDGE SENTENCE PAIRS (MEAN AHR), STATIS (CRISP) AND FUSE (FUZZY) SIMILARITY MEASURES

Hedge Sentence Pairs	AHR	STASIS	FUSE
HSP 1	0.031250	0.22422	0.19360
HSP 2	0.018750	0.53525	0.60376
HSP 3	0.037500	0.31055	0.32459
HSP 4	0.445625	0.33328	0.66473
HSP 5	0.455625	0.62723	0.86166
HSP 6	0.556250	0.66715	0.92492
HSP 7	0.614375	0.69681	0.96199
HSP 8	0.610625	0.73844	0.82998
HSP 9	0.753125	0.85165	0.90680
HSP 10	0.763750	0.87838	0.90734
HSP 11	0.813750	0.92209	0.97473
HSP 12	0.785000	0.76266	0.92203
HSP 13	0.885000	0.46925	0.65697
HSP 14	0.938125	0.88878	0.89207
HSP 15	0.940625	0.90334	0.92420
HSP 16	0.914375	0.99633	0.99243

TABLE IIIB. DISTRIBUTION OF THE HUMAN RATINGS OF HEDGE SENTENCE PAIRS

Hedge Sentence Pairs	Min	Max	Mean	Median
HSP 1	0	5	0.3125	0
HSP 2	0	3	0.1875	0
HSP 3	0	6	0.375	0
HSP 4	2	6	4.45625	4.85
HSP 5	2	6	4.55625	4.75
HSP 6	1	7.2	5.5625	6.25
HSP 7	2	9	6.14375	6.1
HSP 8	3	7.4	6.10625	6.5
HSP 9	4	9.5	7.53125	7.9
HSP 10	4	8.8	7.6375	8.1
HSP 11	5	9	8.1375	8.5
HSP 12	2	9.2	7.85	8.7
HSP 13	6	9.7	8.85	9.25
HSP 14	8	10	9.38125	9.5
HSP 15	7	10	9.40625	9.5
HSP 16	3	10	9.14375	9.8

hedge words chosen (Hedge Words). It gives a (Total Average), which is the average of each hedge row, that is then scaled between [-1, +1] (Scaled) to match the rest of the values scaling in the fuzzy dictionary, ordered from low to high. On examimning the results it can be seen that *Very Fair* is more positively intensified than *Fair*, and the results indicate this closely i.e. *Fair*= 0.085 and *Very Fair*= 0.285. The same applies to *Mild*= -0.2387 and *Very Mild*= 0.285; thus the hedge *Very* positively intensifies a fuzzy word between the ranges of [0.0462,..,0.37]. An example of the affect of negative intensity is the hedge word *Below*, with *Below Fair* = -0.2173 and *Below Mild* = -0.5411, thus *Below* negatively intensifies a fuzzy word between the range of [-0.541,..,-0.2173].

B) Hedge Sentence Pairs results

Table IIIA shows the average human ratings (AHR) obtained from the 16 participants who rated the HSPs. The 16 participants were different from those who had taken part in the Hedge Intensity Experiments outlined in Section III(B); all of whom were native English speakers above the age of 18. Sentence similarity measurements are shown for FUSE and for comparison the similarity is also shown for the measure STASIS which does not incorporate any fuzzy words. Table IIIB shows the distribution of the human ratings showing the Minimum, Maximum, Mean and Median values for each of the 16 sentence pairs.

Table IVA shows one example of a hedge sentence pair (HSP) with average human rating (AHR= 0.8850) taken from Table IIIA. The hedges used in this example are *around* and *very*. The fuzzy words in the sentence pairs are *almost* and *rather* belonging to the category *Level of Membership*, and *dreadfully* and *unpleasantly* belonging to the category *Worth*. STASIS ignores all fuzzy and hedge words and therefore similarity is low (STASIS=0.46925), FUSE on the other hand caters for both fuzzy words and hedge words, therefore has a higher similarity rating (FUSE=0.65697) which is closer to the AHR. This goes to show that fuzzy words and hedge words play an important role in the similarity rating of a short text. On the other hand, Table IVB which relates to HSP12 shows that STASIS (0.76266) has a closer rating to the AHR (0.785000) then FUSE (0.92203). This is likely to be due to the human sample size being relatively small [13] and/or the variations of WordNet used in STASIS and FUSE, as WordNet is constantly being updated.

Looking at the Inter-Rater Correlation in Table V, FUSE gave a higher correlation to Average Human Ratings, with 0.803, compared to STASIS with Average Human Ratings at 0.796. Although the correlation difference was not significant, it is still an improvement over STASIS, which shows that fuzzy hedge intensity does play an important role in sentence similarity. This small improvement can be attributed to 1) the fact that only twelve hedge words were modelled, 2) the coverage of the hedge words in the HSP dataset was limited and 3) the number of human raters was only 16 – acceptable in the NLP community but on the low end of the scale where 32 participants is typically recommended.

Conduction of an Inter-Rater Correlation produces some positive results as can be seen in Table V, with FUSE=0.886 as opposed to STASIS=0.796.

Cicchetti gives the following guidelines for intra-class correlation coefficient agreement measures [25]:

- Less than 0.40 - Poor.
- Between 0.40 and 0.59 - Fair.
- Between 0.60 and 0.74 - Good.
- Between 0. 75 and 1.00 – Excellent

Each of the algorithms STASIS and FUSE is compared against the Average Human Ratings (AHR). Looking at the AHR which is referred to as (a) in this instance, for each of the algorithms it can be seen that in Table VI for STASIS (a= 0.865) and in Table VII for FUSE (a= 0.867) with a confidence interval of 95%. Based on Cicchetti's guidelines, it can be concluded that the intra-class correlation coefficient is deemed as excellent for both datasets.

TABLE IVA. A GOOD EXAMPLE OF HSP

Hedge Sentence Pairs	Sentence 1	Sentence 2	AHR	STASIS	FUSE
HSP 13	And he laughed around almost dreadfully	And he laughed very rather unpleasantly	0.885000	0.46925	0.65697

TABLE IVB. A BAD EXAMPLE OF HSP

Hedge Sentence Pairs	Sentence 1	Sentence 2	AHR	STASIS	FUSE
HSP 12	What s the fine roughly pensionable man	What's the roughly good old man	0.785000	0.76266	0.92203

TABLE V. INTER-RATER CORRELATION OF FUSE & STASIS

Inter-Rater Correlation Matrix			
	STASIS	FUSE	AHR
STASIS	1.000	0.886	0.796
FUSE	0.886	1.000	0.803
AHR	0.796	0.803	1.000

TABLE VI. INTRA-CLASS CORRELATION COEFFICIENT FOR STASIS

Intra-class Correlation Coefficient			
	Intra-class Correlation	95% Confidence Interval	
		Lower Bound	Upper Bound
Single Measures	.762	0.442	0.910
Average Measures	.865	0.613	0.953

TABLE VII. INTRA-CLASS CORRELATION COEFFICIENT FOR FUSE

Intra-class Correlation Coefficient			
	Intra-class Correlation	95% Confidence Interval	
		Lower Bound	Upper Bound
Single Measures	.766	0.450	0.911
Average Measures	.867	0.620	0.954

V. CONCLUSION AND FURTHER WORK

This paper has presented a study on the application of linguistic hedges within fuzzy semantic similarity measures. This has involved first obtaining human intensity ratings of a small selection of hedges to fuzzy words. These hedges were then modelled using Type-II interval fuzzy sets for inclusion in the FUSE fuzzy dictionary. A set of 16 hedge sentence pairs were constructed using the modelled hedges and 16 participants rated their similarity. Although there was minor improvement on the similarity measurement correlation between average human ratings and the fuzzy measure FUSE, it was not significant. This is mainly due to the number of hedges modelled and the number of participants involved in rating the hedge sentence pairs. However even with this small sample, it can be seen that linguistically modelled hedges have a positive effect on sentence similarity. Current work consists of, but is not limited to, expanding the hedge sentence pairs and also expanding the sample size to cater for more human ratings. A future experiment will investigate the impact of hedges on the degree of intensification of a sentence, by determining the fuzzy similarity of pairs of sentences, first with hedges and then without, and comparing both results to the average human rating of each variation. This would allow a greater evaluation of the impact of hedge words applied to individual fuzzy words beyond this paper by looking at how a human interprets the hedge words in the context of a sentence.

Current work is incorporating FUSE into dialogue systems, which will allow a wider range of natural language dialogue to be explored and tested in real-world dialogue utterance exchanges.

ACKNOWLEDGMENT

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) under reference UID/CEC/50021/2019, grant SFRH/BSAB/136312/2018 and project LISBOA-01-0145-FEDER-031474

REFERENCES

[1] Zadeh, L.A., "Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges", Journal of Cybernetics, 2:3, 1972, pp. 4-34, DOI: 10.1080/01969727208542910

[2] Kerre E.E., De Cock M., "Linguistic Modifiers: An Overview. In: Chen G., Ying M., Cai KY. (eds) Fuzzy Logic and Soft Computing." The International Series on Asian Studies in Computer and Information Science, vol 6, 1999, Springer, Boston, MA

[3] Shi, H., Ward, R., Kharma, N., "Expanding the definitions of linguistic hedges", Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), 2001

[4] Novak, V., Fuzzy logic in natural language processing, in Proc. IEEE Int. Conf. Fuzzy Systems Conf.erence, 2017.

[5] Le,V., Tran, D., Extending fuzzy logics with many hedges, Fuzzy Sets and Systems, Vol345, 2018, pp.126-138, https://doi.org/10.1016/j.fss.2018.01.014

[6] Wang, H. Xu, Z. Zeng, X., Linguistic terms with weakened hedges: A model for qualitative decision making under uncertainty, Information Sciences, Volumes 433–434, 2018, pp. 37-54.

[7] Wang, H. Xu, Z. Zeng, X, Huchang, L., Consistency measures of linguistic preference relations with hedges, 2018, IEEE Transactions on Fuzzy Systems, DOI: 10.1109/TFUZZ.2018.2856107.

[8] Hameed, I., Enhanced fuzzy system for student's academic evaluation using linguistic hedges,, 2017, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), DOI: 10.1109/FUZZ-IEEE.2017.8015462

[9] Djatna, T., Luthfiyanti, R., Abbas, A., "Intuitionistic fuzzy hedges modeling for supplier selection of responsive agroindustrial multi products supply chains in small and medium enterprises", 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), DOI: 10.1109/ICACSIS.2017.8355042,

[10] Borkar, P., Sarode, M.V. and Malik, L.G., Int. J. Fuzzy Syst., 2016, 18: 379. https://doi-org.ezproxy.mmu.ac.uk/10.1007/s40815-015-0069-5

[11] Chandran, D., "The development of a fuzzy semantic sentence similarity measure," Doctorate of Philosophy, School of Computing, Maths and Digital Technology, Manchester Metropolitan University (MMU), 2013.

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

[13] O'Shea, J.D., Bandar, Z.A. and Crockett, K. "A new benchmark dataset with production methodology for short text semantic similarity algorithms," *ACM Transactions on Speech and Language Processing (TSLP)*, vol. 10, no. 4, 2013, p. 19.

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

[15] Mendel, J.M. and Wu, D., "Determining interval type-2 fuzzy set models for words using data collected from one subject: Person FOUs", 2014 IEEE International Conference on Fuzzy Systems, pp.768 – 775, 2014.

[16] Mendel, J.M., "Type-2 Fuzzy Sets as Well as Computing with Words", Published in: IEEE Computational Intelligence Magazine (Volume: 14 , Issue: 1 , Feb. 2019), DOI: 10.1109/MCI.2018.2881646,

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

[18] Li,Y., McLean, D., Bandar, Z.A., O'Shea, J.D. and Crockett, K. "Sentence similarity based on semantic nets and corpus statistics"*, IEEE transactions on knowledge and data engineering*, vol. 18, no. 8, 2006, pp. 1138-1150.

[19] Princeton University, "About Wordnet." Retrieved 13 June 2014 from http://wordnet.princeton.edu/

[20] Negnevitsky, M., Artificial Intelligence: A Guide to Intelligent Systems. 2 ed., 2005, p. 435. England: Pearson Education Limited.

[21] Banks, W., (2003) "Linguistic Variables: Clear Thinking with Fuzzy Logic." Retrieved 30 May 2016 from http://www.phaedsys.co.uk/principals/bytecraft/bytecraftdata/LinguisticVariables.pdf

[22] Zadeh, L.A., "The Concept of a Linguistic Variable and its Applications to Approximate Reasoning", Information Sciences, Vol. 8,1975, pp 199-249 (Part I), pp 301-357 (Part 11), Vol. 9, 1975, pp 43-80 (Part In)

[23] Cox, E. "The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems" vol. 1,. AP Professional, 1994

[24] Feilong, L. and Mendel, L.M., "Encoding words into interval type-2 fuzzy sets using an interval approach." IEEE Transactions on Fuzzy Systems, vol. 16, no. 6, 2008, pp.1503-1521.

[25] Cicchetti, D.V., "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology." Psychological Assessment, vol. 6, no. 4, 1994, p. 284.