DOI: https://doi.org/10.1109/FUZZ52849.2023.10309737

Publisher: IEEE

Version: Accepted Version

An Investigation into Fuzzy Negation in Semantic Similarity Measures

Naeemeh Adel
Department of Computing and Mathematics,
Manchester Metropolitan University,
Chester Street,
Manchester, United Kingdom

Keeley Crockett
Department of Computing and Mathematics,
Manchester Metropolitan University,
Chester Street,
Manchester, United Kingdom

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

***Abstract*—Machine computation of semantic similarity between short texts aims to approximate human measurements of similarity, often influenced by context, domain knowledge, and life experiences. Logical negation in natural language plays an important role as it can change the polarity of meaning within a sentence, yet it is a complex problem for semantic similarity measures to identify and measure. This paper investigates the impact of logical negation on determining fuzzy semantic similarity between short texts containing fuzzy words. A methodology is proposed to interpret the implications of a negation word on a fuzzy word within the context of a user utterance. Three known fuzzy logical not operators proposed by Zadeh, Yager and Sugeno are incorporated into a fuzzy semantic similarity measure called FUSE. Experiments are conducted on a sample dataset of short text inputs captured through human engagement with a dialogue system. Results show that Yager's weighted operator is the most suitable for achieving a matching threshold of 90.47% accuracy. This finding has significant implications for the field of semantic similarity measures. It provides a more accurate way to measure the similarity of short texts that contain fuzzy words combined with logical negation. Whilst validation of the approach on more substantial datasets is required, this study contributes to a better understanding of how to account for logical negation in fuzzy semantic similarity measures and provides a valuable methodology for future research in this area.**

Keywords— fuzzy semantic similarity, logical negation, FUSE, natural language processing

I. INTRODUCTION

Negation is a complex concept in human language which has a rich history that was first collated by Horn [1] as stemming from Aristotle. The Cambridge Dictionary describes negation as "*the action of causing something to not exist or to have no effect*" [2] and "*the exact opposite of something, or a complete lack of it*" [2]. Morante and Blanco [3] summarise that from a linguistics point of view, negation has both a scope and focus. The scope is defined as part of the meaning in a language structure, whereas the focus is "*the part of the scope that is most prominently or explicitly negated*" [2]. A summary of recent advances of negation in linguistics can be found here [3, 4, 5].

Logical negation in natural language plays an important role as it can often change the polarity of a sentence from a positive one to a negative one and vice versa [6]. For example, given the two sentences S1 and S2:

S1: [This food was really worth waiting for]

S2: [This food was really <u>not</u> worth waiting for]

Both S1 and S2 have the same number of words in the same order with the only difference being the presence of the word '*not*' in S2 before the word *worth* which completely changes the polarity of S2 from a positive experience to a negative one. On the other hand, referring to the following two sentences S3 and S4:

S3: [The season finale was predictable]

S4: [The season finale was <u>un</u>predictable]

Here, S3 has a negative meaning in the context of a show having a predictable ending, but S4 (again same number of words in the same order) with the presence of the prefix '*un*' takes this negative experience and turns it into something positive in the context, that the season finale of the show was actually unpredictable, meaning it was exciting and amusing. The use of a fuzzy complement operator has not been considered specifically within the field of Short Text Semantic Similarity Measures (STSM), yet it is an important concept as illustrated by the two examples ([S1, S2] and [S3, S4]).

A known challenge in the field of semantic sentence similarity is the inability of traditional STSM to capture the meaning of fuzzy words. A fuzzy word can be defined as *a natural language word with a subjective meaning* [7]; e.g., *huge*, *small*, *hot* and *cold*, that is characteristically used in everyday human natural language dialogue [7]. Fuzzy words are often ambiguous in meaning since they are based on an individual's perception of a specific context. To date, two fuzzy STSM have been developed, FAST [8] and FUSE [7] but neither has attempted to address the measurement of negation in associated fuzzy words when computing the similarity of short texts, nor assess the impact on the correlation of the similarity measurement with respect to human ratings.

The impact of negation on the measurement of the similarity of short texts was only fully understood after the integration of the FUSE algorithm into a dialogue system (DS) [9]. In [9], FUSE was incorporated into a simple linear DS designed to provide consumer feedback on a cafe visit. In this study, 32 participants answered nine free-form text questions giving a total of 288 human responses. These responses were compared with the six prototypical sentences for each of the nine fuzzy categories generating a dataset of 1728 sentence pairs. Whilst FUSE generally performed well, obtaining a true positive matching rate of 87.5% of human utterances, it was found that fuzzy matching of user utterances that contained negative connotations with dialogue system prototypical responses were poor. Further empirical experiments (in a forthcoming publication) with two other dialogue systems provided evidence that the implication of negation words did not impact the fuzzy semantic similarity score. The motivation of this study is to investigate how traditional fuzzy negations can be applied in the context of natural language semantic understanding when comparing STSM.

In this paper, we present the results of a preliminary investigation into the use of fuzzy complement operators within FUSE. The key contributions of this work are a methodology to measure the effect of fuzzy semantic similarity of a negation word on a connecting fuzzy word within a human utterance when computing fuzzy short text semantic similarity. A further contribution is the inclusion of the method within the FUSE algorithm where it is verified on human utterances captured within a dialogue system.

The rest of this paper is organised as follows: Section II provides background on semantic similarity measures and a brief review of relevant fuzzy negation operators. Section III defines the problem of interpreting and understanding negation in the context of fuzzy semantic similarity measures. Section IV describes the experimental methodology and empirical results on a dataset of human utterances. Finally, Section V presents the conclusions and future work.

II. BACKGROUND AND RELATED WORKS

A. *Semantic and Fuzzy Semantic Similarity Measures*

Short Text Semantic Similarity Measures (STSM) are used to quantify the degree to which short texts (25 - 30 words in length) are semantically equivalent to each other in correlation to subjective evaluation by humans [10]. Semantic similarity is, therefore, a complex concept with a long history in cognitive psychology and linguistics [11], which can analyse the deep semantic structure of a short text to convey meaning. Semantic similarity methods usually give a ranking or percentage of similarity between texts as opposed to a binary decision [12]. The task of assessing the semantic similarity between short texts has been a central problem in natural language processing, due to its importance in a variety of applications [13].

FUSE is an ontology-based similarity measure that uses Interval Type-2 Fuzzy Sets to model relationships between categories of human perception-based words. Several versions of FUSE (FUSE_1.0 – FUSE_4.0) have been developed, investigating the presence of linguistic hedges, the expansion of fuzzy categories and their use in natural language, and the introduction of the fuzzy influence factor [14, 15, 16]. FUSE has been compared to several state-of-the-art, STSM which do not consider the presence of fuzzy words in the similarity calculation. Results have shown FUSE is able to improve on the limitations of crisp STSM by achieving a higher correlation with the Average Human Rating (AHR) compared to traditional Sentence Similarity Measures (SSM) using several published and gold-standard datasets. The full FUSE algorithm can be found in [16] along with a fuzzy dictionary comprising of nine fuzzy categories (Size/Distance, Temperature, Brightness, Age, Speed, Strength, Frequency, Level of Membership, and Worth) of fuzzy words. In the research presented in this paper, a variation to the FUSE similarity measure is proposed, to assess the impact of negation on fuzzy words when computing similarity. Empirical experiments are then used to investigate a series of fuzzy negation operators on the measurement of similarity.

B. *Logical Not operators*

One of the key challenges when processing human language is negation identification. Sergeeva et al. [17] proposed a feature-heavy long short-term memory network (LSTM) based model for negation scope detection in biomedical texts to examine the dependency on the use of negation connotations on a series of reported medical events within the same text. Pabon et al. [18] also adopted a deep learning approach to negation and uncertainty detection in Spanish clinical texts. Within sentiment analysis [19], negation is often handled through recognition of negation trigger words such as '*not*' that reverses the polarity of words in its vicinity or through a process of classification, to detect positive or negative words or emojis. Pradhan et al. [19] utilises a hybrid polarity shift (negation) detection approach for aspect-based sentiment analysis to investigate viewpoints of customers about similar products which obtained good results compared with the state of the art. Lal and Kamath [20] provided an approach to handle sentiment negation using synsets in the WordNet lexical database. Novak [21] first introduced the concept of Fuzzy Natural Logic (FNL) to produce a model of linguistic semantics which went on to introduce a mechanism to deal with the linguistic vagueness of evaluative expressions [22]. A notable contribution of this work is in the proof-of-concept creation of an ontology of evaluative expressions for computational applications.

Fuzzy negations or fuzzy complements were first introduced by Zadeh [23]. Ferri et al. [24] highlighted the challenge of fuzzy negation in the realm of natural language, human interpretation and common sense, understanding complements may not be plausible. Ferri et al. [24] gave the example of membership functions '*old*' and '*very-old*' within the context of age [0-130], where '*old*' was defined using a triangular membership function within the fuzzy subset of age [72-96] where the age of 84 had a membership grade in '*old*' with a degree of 1. Producing the fuzzy set '*not-old*' would yield a fuzzy set which considers all other ages across the domain, yet in a human language discussion concerning babies, it would be very unlikely a baby would be referred to in a conversation as '*not-old*'. Thus Ferri [24] argues that fuzzy complements contain limited meaning in the part of the domain where the grade of membership is 1.

The original complement operator introduced by Zadeh is defined by taking one minus the membership value at each point, along the truth function, with no additional parameters needed as shown in eq. 1. Since the complement of a fuzzy set is often used as a new fuzzy region in a model, the 'complement' or 'negation' is produced by creating and populating a new fuzzy set [23].

$$\sim\mu_A(x) = (1 - \mu_1) \quad (1)$$

Several methods have also been proposed on how to apply fuzzy complements within fuzzy systems. These include the historically popular Sugeno [25] and Yager [26] complements. The Sugeno complement takes a class parameter that determines the strength of the negation. The Sugeno class is defined in eq. 2 [27]:

$$\sim\mu_A(x) = \frac{1 - \mu_A(x)}{1 + k\mu_A(x)} \quad (2)$$

In this case, the class parameters are in the range [-1, ∞]. When $k = 0$ the Sugeno complement has the desirable property of becoming the standard Zadeh complement [23]. Yager defines an alternative form of the fuzzy complement having a power function as defined in eq. 3 [27]:

$$\sim\mu_A(x) = (1 - \mu_1(x)^k)^{\frac{1}{k}} \quad (3)$$

where the class function k is generally in the range [>0, <5]. The class function performs the standard Zadeh complement (which is found when $k = 1$). The class membership in the Yager complement provides a convenient and flexible method of adjusting the strength of the fuzzy '*not*' operator. For the endpoint conditions of zero and one, the Yager complement, regardless of the class strength parameter, always acts like the standard Zadeh complement [23]. A summary of fuzzy complement measures can be found in [28].

III. PROBLEM DETERMINATION

To test the ability of the FUSE measure to deal with real-world human utterances, a dialogue system (DS) was designed called FUSION_V1. FUSE was incorporated to measure the similarity of human responses to sets of question-specific prototypical sentences, thus alleviating the need for cumbersome pattern-matching approaches [9]. Human participants were invited to take part in an observation scenario and answer a set of questions designed to capture answers relating to each of the nine fuzzy categories. The scenario was based on the participant attending a local cafe and ordering a drink of their choice and sitting down and observing their surroundings for a few minutes. After this, FUSION_V1 DS asked each participant nine questions, one question per fuzzy category, and participants typed their answers into the dialogue system. Each of the nine fuzzy questions had three pairs of prototypical sentences broken down into thresholds of *high*, *medium,* and *low*. Fig. 1 shows an example of one of the fuzzy questions relating to the cafe

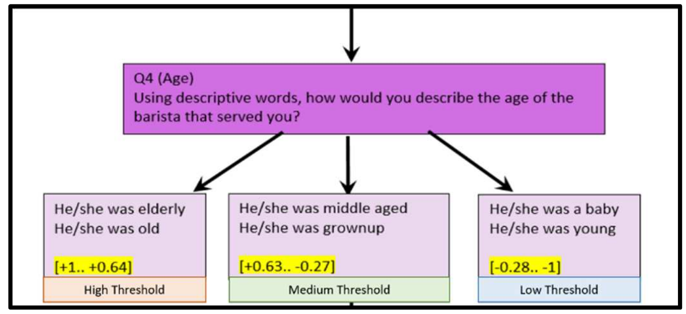

Fig. 1 Sample Question and Associated Prototypical Sentences

TABLE I. EXAMPLE OF INCORRECT SIMILARITY MEASUREMENT IN FUSE

	Human Responses	Prototypical Sentences	FUSE without NOT
High	The light level of the cafe is not bright	The cafe was bright	0.890023465
	The light level of the cafe is not bright	The cafe was light	0.894193583
Medium	The light level of the cafe is not bright	The cafe was twinkling	0.774757829
	The light level of the cafe is not bright	The cafe was luminous	0.874176525
Low	The light level of the cafe is not bright	The cafe was moonlit	0.731469762
	The light level of the cafe is not bright	The cafe was lightless	0.619931610

scenario dialogue system and the three thresholds, with two sentence pairs in each threshold. The values that determined the threshold differ per category as explained in [9]. [9] further describes how the participant utterances were collected and run with both FUSE and a traditional STSM known as STASIS [29] embedded within FUSION_V1. During analysis of the results, it was noted that both FUSE and STASIS did not manage to give a correct similarity measurement with sentences that presented logical negation values such as '*not*' when compared to human ratings. Table I shows one of the responses containing the word '*not*' that did not return a similarity to the correct threshold. Originally the sentence '*The light level of the cafe is not bright*' which was a response from a participant under the FUSION_V1 experiment [9], scored the highest similarity rating with the sentence '*The cafe was light*'. Table I shows that this sentence matched the *high* threshold for this category. However, due to the presence of the negation word '*not*' in the participants' sentence, it actually means the cafe is not bright and optimal results would be to score similarity with the *low* threshold sentences (*The cafe was moonlit* or *The cafe was lightless*).

To address the presence of logical negation in fuzzy sentences it was required to formulate a methodology to correctly interpret the implications of a negation word on a fuzzy word within the context of a user utterance. Therefore, a preliminary experiment was conducted on the aforementioned sample sentence '*The light level of the cafe is not bright*' and the six prototypical sentences assigned to this question as shown in Table II to investigate the inclusion of the theoretical fuzzy complement operators using the FUSE similarity calculation. Table II shows the participant responses in the second column (*Sentence 1*), the six threshold sentences for the Brightness category in column three (*Sentence 2*), and the (*Threshold*) column indicates which threshold each sentence within the *Sentence 2* column belongs to (*High*, *Medium* or *Low*). To determine which method provided the best results in terms of the highest correlation to human ratings a short experiment was conducted to test the three measures of Zadeh [23], Sugeno [25], and Yager [26] as defined in Section II. For example, the class membership in the Yager complement provides a convenient and flexible method of adjusting the strength (class parameters) of the fuzzy '*not*' operator. Klir [27] suggests using the following class strengths for Sugeno [k = 10, k = 2, k = 0, k = -0.5, k = -0.9] and the following class strengths Yager [k = 0.5, k = 1, k = 2, k = 5] to assess the impact of the application of different strengths of negation with the class parameters being in the range of [-1, ∞].

IV. Empirical Investigations

This section explores three logical negation operators proposed by Zadeh [23], Sugeno [25], and Yager [26] which are embedded into the FUSE algorithm. A preliminary experiment is used to explore these operators and investigate a range of weights (if applicable) to determine the effects on the measurement of short text similarity. The most promising operator is then selected, embedded within the FUSE algorithm and evaluated on a real-world Fuzzy Not Dataset (FND) which has been formulated via human engagement using a fuzzy dialogue system.

A. *Preliminary experiment*

To carry out this preliminary experiment a dataset of six sentences was created using the sentence pairs shown in Table II. The optimum results for this particular sentence '*The light level of the cafe is not bright*' would be for the similarity to fall in the *Low* threshold and have a similarity with the prototypical sentence '*The cafe was lightless*'. Table III shows the results of the experiment conducted on the six sentence pairs from Table II presented on a scale of [0, 1]. The column labelled *Original* shows the similarity levels returned for each sentence pair in Table II using FUSE without running any negation parameters. The values highlighted in yellow for each column represent the highest similarity score for that class parameter (Zadeh, Sugeno, Yager) and the prototypical sentence that it matched against. The original defuzzified value for the word *bright* is (0.57) in the FUSE fuzzy dictionary. Using the Yager class (k = 0.5) the word '*not bright*' is given a measure of (-0.8799) and the highest similarity is achieved with SP6, as seen in Table III, highlighted in pink. This was the prototypical sentence that was the ideal sentence to be matched against with a similarity rating of (0.9277). Therefore, the Yager class with a strength of (k = 0.5) was used with FUSE to carry out the next phase of the experiment involving human responses containing the logical not operator. The authors acknowledge that further extensive experimentation would be required on a wide range of datasets to validate whether (k = 0.5) was the sub-optimal generalised value.

B. *Formulation of a Fuzzy Not Dataset*

Datasets for the evaluation of fuzzy semantic similarity measures are limited [9]. Since no data sets specifically containing significant proportions of logical negation phrases or words such as '*not*' exist, the proposed method could not be tested effectively with FUSE. Therefore, a second dialogue system referred to as FUSION_V2 consisting of 18 fuzzy questions, two questions representing each of the nine fuzzy categories (Size/Distance, Temperature, Brightness, Age, Speed, Strength, Frequency, Level of Membership, Worth) [16] was designed. Due to the COVID-19 global pandemic and the implications of shutting down many universities and offices, people were forced to work from home (WFH) on very short notice. This in turn involved certain alterations that had to be made to people's homes and working habits to allow these new adjustments to their working conditions. Therefore, the FUSION_V2 DS was designed around this scenario to ask participants questions relating to their WFH conditions. A detailed description of FUSION_V2, in-depth analysis and results are the subject of a forthcoming journal publication. To effectively test the functionality of the fuzzy logical negation in the FUSE algorithm, the participant results from both FUSION_V1 and FUSION_V2 that contained the word '*not*', followed immediately by a fuzzy word present in the fuzzy dictionary [16] were collated comprising of 21 human responses from both the Cafe scenario and the WFH scenario, as shown in Table IV. Each of the 21 human responses was aligned with the six prototypical sentences for that given question giving a total of 126 sentence pairs referred to as the Fuzzy Not Dataset (FND). To effectively test the logical negation implemented in the FUSE algorithm, FND was run firstly

TABLE II. Brightness Category Sentence Thresholds

	Sentence 1	Sentence 2	Threshold
SP 1	The light level of the cafe is not bright	The cafe was bright	High
SP 2	The light level of the cafe is not bright	The cafe was light	
SP 3	The light level of the cafe is not bright	The cafe was twinkling	Medium
SP 4	The light level of the cafe is not bright	The cafe was luminous	
SP 5	The light level of the cafe is not bright	The cafe was moonlit	Low
SP 6	The light level of the cafe is not bright	The cafe was lightless	

TABLE III. Preliminary Experimental results of fuzzy negation operators

Sentence Pair	Original	Zadeh	Sugeno					Yager			
		$1-\mu_1$	k = 10	k = 2	k = 0	k = -0.5	k = -0.9	k = 0.5	k = 1	k = 2	k = 5
SP 1	0.9770	0.8024	0.7215	0.7811	0.8024	0.9531	0.9643	0.7215	0.8024	0.9643	0.9335
SP 2	0.9778	0.7787	0.7263	0.7647	0.7787	0.9335	0.9453	0.7263	0.7787	0.9453	0.9138
SP 3	0.8772	0.8759	0.6877	0.7740	0.8759	0.9257	0.8240	0.6877	0.8759	0.8240	0.7740
SP 4	0.9660	0.8507	0.7309	0.8069	0.8507	0.9912	0.9373	0.6834	0.8507	0.9373	0.8952
SP 5	0.8363	0.9410	0.8951	0.9770	0.9410	0.8459	0.7528	0.8951	0.9410	0.7528	0.7128
SP 6	0.6599	0.7552	0.8277	0.9231	0.7552	0.6502	0.5882	0.9277	0.7552	0.5882	0.5523

with FUSE without logical negation, referred to as (FUSE without NOT) and then using FUSE with logical negation, referred to as (FUSE with NOT), using the Yager class with a strength of (k = 0.5) [26]. To compare fuzzy sentence similarity measures, FND was also run with STASIS [29] which does not cater for any fuzzy words or logical negation.

To calculate the fuzzy logical negation values, present in the 21 human responses of FND, for any fuzzy words that had the word '*not*' present immediately before them, the defuzzified value for the fuzzy word present in the human response was obtained from the FUSE fuzzy dictionary [16] and using the Yager class with a strength of (k = 0.5) the fuzzy logical negation value was obtained. For example, the word '*dazzling*' present in HR 5, as shown in Table IV, (had a rating of 0.6) belonging to the Brightness category, and the user response of '*not dazzling*' (will have a rating of -0.8984).

C. *Experimental Results and Discussion*

FND was run with STASIS, FUSE without NOT and FUSE with NOT for the 126 sentence pairs. The results showed that from the 21 human utterances used in FND, FUSE with NOT had a total of 19 human responses returning a similarity that matched with the correct threshold, giving an accuracy of 90.47%. Tables IV–VII show examples of four human responses with the columns *Human Responses* representing the sentence given by the human participant, *Prototypical Sentence* column which represents the six prototypical sentences based on the question asked by the dialogue system, and the three columns of STASIS, FUSE without NOT and FUSE with NOT representing the sentence similarity measures which are being used to return a similarity score per sentence pair. Where there is a column highlighted in red, it is showing the original threshold that each sentence pair scored the highest similarity when run with STASIS and FUSE without NOT. Where there is a column highlighted in blue it is showing the correct threshold that each sentence pair scored the highest similarity when run with FUSE with NOT. Looking at Table IV in more detail, the human response states that '*Lighting's relatively good, not bright*'. The presence of the word '*not*' before the fuzzy word '*bright*' indicates that the lighting is in fact relatively dim. STASIS and FUSE without NOT returned a similarity which fell in the *high* threshold due to the presence of the word '*bright*' in the human response being matched to the word '*bright*' in the prototypical sentence, since STASIS and FUSE without NOT do not cater

TABLE VIII. 21 HUMAN RESPONSES FROM CAFE SCENARIO AND WFH SCENARIO

Fuzzy Category	Human Responses (HR)	
Size/Distance	HR 1	In my experience I would say that the queue was **not long** as it is usually
Temperature	HR 2	The temperature of the room was a normal temperature, it was **not hot**.
Brightness	HR 3	I have no glare on my monitor as it is **not bright** with the blind shut
	HR 4	Lighting's relatively good, **not bright**.
	HR 5	Ambient, just right, but **not dazzling**.
	HR 6	The lighting is alright it's **not bright** nor dim. Sometimes, I open the blinds during the day and get some sunlight.
	HR 7	**Not bright** – just right !
	HR 8	It is ok, **not dark**
Strength	HR 9	I am **not healthy** as I tend to stay at my desk longer at home than at work and I need more exercise.
	HR 10	**Not fit** - medium
	HR 11	She is **not healthy**
	HR 12	**Not healthy**
	HR 13	The physical state of the last person I saw was **not fit**
	HR 14	I would describe them as lean and **not strong**.
	HR 15	He is slim built so I would suggest he is physically **not strong** at all
Worth	HR 16	**Not satisfied** at all but don't have much of a choice.
	HR 17	**Not satisfied** at all !
	HR 18	I am **not satisfied** with my working conditions when working from home.
	HR 19	I would say I am **not satisfied** with the furniture I am currently using
	HR 20	I am **not satisfied** with the furniture I use
	HR 21	I am **not satisfied** with the work furniture that I am using.

TABLE IV. HR 4 RESULTS (YAGER CLASS K = 0.5)

	Human Responses	Prototypical Sentences	STASIS	FUSE without NOT	FUSE with NOT
High	Lighting's relatively good, **not bright**.	The lighting is light	0.887766753	0.915166772	0.818589283
	Lighting's relatively good, **not bright**.	The lighting is bright	0.920953436	0.920175238	0.801111894
Medium	Lighting's relatively good, **not bright**.	The lighting is sunlit	0.814019904	0.856087858	0.676223709
	Lighting's relatively good, **not bright**.	The lighting is beaming	0.716587213	0.895415332	0.767506811
Low	Lighting's relatively good, **not bright**.	The lighting is moonlit	0.830699358	0.832311889	0.887873420
	Lighting's relatively good, **not bright**.	The lighting is lightless	0.814019904	0.752680766	0.939218108

TABLE V. HR 5 RESULTS (YAGER CLASS K = 0.5)

	Human Responses	Prototypical Sentences	STASIS	FUSE without NOT	FUSE with NOT
High	Ambient, just right, but **not dazzling**.	The lighting is light	0.255154759	0.571623908	0.349349722
	Ambient, just right, but **not dazzling**.	The lighting is bright	0.206232149	0.615625720	0.375623444
Medium	Ambient, just right, but **not dazzling**.	The lighting is sunlit	0.187496277	0.815848797	0.333772447
	Ambient, just right, but **not dazzling**.	The lighting is beaming	0.160402960	0.598765646	0.417374891
Low	Ambient, just right, but **not dazzling**.	The lighting is moonlit	0.193256982	0.407516464	0.668331149
	Ambient, just right, but **not dazzling**.	The lighting is lightless	0.187496277	0.349790816	0.831041135

TABLE VI. HR 13 RESULTS (YAGER CLASS K = 0.5)

	Human Responses	Prototypical Sentences	STASIS	FUSE without NOT	FUSE with NOT
High	The physical state of the last person I saw was **not fit**.	The last person I saw looked tough	0.852803783	0.852803783	0.687222391
	The physical state of the last person I saw was **not fit**.	The last person I saw looked strong	0.853394955	0.869518089	0.670254941
Medium	The physical state of the last person I saw was **not fit**.	The last person I saw looked athletic	0.827082256	0.915760160	0.710320025
	The physical state of the last person I saw was **not fit**.	The last person I saw looked energetic	0.820027993	0.915029884	0.706734410
Low	The physical state of the last person I saw was **not fit**.	The last person I saw looked weak	0.834685144	0.792903442	0.914659837
	The physical state of the last person I saw was **not fit**.	The last person I saw looked delicate	0.832520731	0.806581158	0.931940764

TABLE VII. HR 16 RESULTS (YAGER CLASS K = 0.5)

	Human Responses	Prototypical Sentences	STASIS	FUSE without NOT	FUSE with NOT
High	**Not satisfied** at all but don't have much of a choice.	My current working conditions are wonderful	0.296424212	0.305397824	0.314357468
	Not satisfied at all but don't have much of a choice.	My current working conditions are amazing	0.296047125	0.302517840	0.306243328
Medium	**Not satisfied** at all but don't have much of a choice.	My current working conditions are average	0.297767144	0.360815425	0.339079320
	Not satisfied at all but don't have much of a choice.	My current working conditions are alright	0.277505193	0.403615169	0.455770183
Low	**Not satisfied** at all but don't have much of a choice.	My current working conditions are ok	0.266074731	0.354030482	0.514078292
	Not satisfied at all but don't have much of a choice.	My current working conditions are unbearable	0.288456997	0.288468001	0.673560005

for the presence of the word '*not*' which would inverse the meaning of '*bright*'. On the other hand, it can be seen that the FUSE with NOT column returned a similarity which fell in the *low* threshold matching the word '*not bright*' to '*lightless*'. Two of the human responses did not fall in the desired threshold as shown in Tables VIII and IX. Looking at Table IX in more detail, the human response '*The lighting is alright it's not bright nor dim. Sometimes, I open the blinds during the day and get some sunlight.*' uses a lot of descriptive words in a combination that would potentially fall in the middle threshold, coupled with the presence of the logical operator with the word '*not bright*' causing confusion for the algorithm. This could be the reason why it did not match with the correct threshold even when run with FUSE with NOT.

V. Conclusion and Future Work

The work presented in this paper has shown that firstly the presence of logical negation in fuzzy sentences poses a challenge for fuzzy SSM in correctly interpreting the implications of a negation word in user utterances. To address this issue, the paper proposed a methodology to correctly interpret the effects of a negation word on a fuzzy word in the context of a user utterance. A preliminary experiment was conducted to explore three logical negation operators proposed by Zadeh, Sugeno, and Yager, which were embedded into the FUSE algorithm to determine the effects on the measurement of short text similarity. The Yager class with a strength of ($k = 0.5$) was found to be the most promising operator and was used with FUSE to carry out an evaluation on a real-world Fuzzy Not Dataset. It must be reiterated that extensive testing should be made before deciding the most appropriate measure and the Yager class with a strength of ($k = 0.5$) was only chosen to proceed with the experiments as a proof of concept. To validate the approach, further rigorous experimentation needs to be undertaken on datasets that specifically have a good representative sample of fuzzy negation words. Secondly, optimisation of k, (or other class parameters) should be determined and tested for generalisation. This study demonstrates the potential of FUSE for use in real-world applications and highlights the importance of considering the effects of logical negation in fuzzy sentences.

Further work in this area can include expanding the FUSE algorithm to handle more complex linguistic structures beyond simple negation. At present FUSE addresses '*not*' only when it is immediately present before a fuzzy word (i.e., *not bright*). Further work could involve catering for '*not*' and similar negation values anywhere in the sentence and not just directly before a fuzzy word (i.e., I do *not* want a *large* drink) where *large* is the fuzzy word in the sentence. Datasets used by the NLP community often do not have sufficient logical operators combined with fuzzy words to allow for rigorous testing of a fuzzy SSM. Therefore, one of the challenges would be specific datasets that may need to be created or curated from existing ones. Future research can investigate the application of FUSE in various domains beyond dialogue systems, such as document retrieval and sentiment analysis. Another potential avenue of exploration is the incorporation of FUSE into machine learning models to improve their performance on tasks that involve fuzzy matching. Furthermore, it would be interesting to investigate the performance of FUSE in multilingual scenarios, as different languages may have different rules for handling fuzzy concepts. Overall, the results of this study suggest that FUSE can be a valuable tool for measuring similarity in human utterances and has the potential to be a useful addition to natural language processing applications.

TABLE VIII. HR2 Results (Yager Class k = 0.5)

	Human Responses	Prototypical Sentences	STASIS	FUSE without NOT	FUSE with NOT
High	the temperature of the room was a normal temperature, it was **not hot.**	It was roasting	0.376324421	0.400456109	0.266903066
	the temperature of the room was a normal temperature, it was **not hot.**	It was boiling	0.363889817	0.280049565	0.295828277
Medium	the temperature of the room was a normal temperature, it was **not hot.**	It was mild	0.520898409	0.467958169	0.649631360
	the temperature of the room was a normal temperature, it was **not hot.**	It was frigid	0.586031728	0.567894349	0.447271789
Low	the temperature of the room was a normal temperature, it was **not hot.**	It was freezing	0.360770211	0.695448553	0.792048644
	the temperature of the room was a normal temperature, it was **not hot.**	It was chilly	0.619428075	0.627963652	0.759475246

TABLE IX. HR6 Results (Yager Class k = 0.5)

	Human Responses	Prototypical Sentences	STASIS	FUSE without NOT	FUSE with NOT
High	The lighting is alright it's **not bright** nor dim. Sometimes, I open the blinds during the day and get some sunlight.	The lighting is light	0.699220766	0.682377837	0.640347381
	The lighting is alright it's **not bright** nor dim. Sometimes, I open the blinds during the day and get some sunlight.	The lighting is bright	0.584186799	0.565965965	0.502968742
Medium	The lighting is alright it's **not bright** nor dim. Sometimes, I open the blinds during the day and get some sunlight.	The lighting is sunlit	0.492281749	0.516942114	0.378868933
	The lighting is alright it's **not bright** nor dim. Sometimes, I open the blinds during the day and get some sunlight.	The lighting is beaming	0.472361815	0.549051419	0.468086142
Low	The lighting is alright it's **not bright** nor dim. Sometimes, I open the blinds during the day and get some sunlight.	The lighting is moonlit	0.511691266	0.472979756	0.597519118
	The lighting is alright it's **not bright** nor dim. Sometimes, I open the blinds during the day and get some sunlight.	The lighting is lightless	0.492281749	0.404024173	0.669238218

VI. Acknowledgement

This work was in part supported by Fundação para a Ciência e a Tecnologia (FCT), through Portuguese national funds Ref. UIDB/50021/2020, and Agência Nacional de Inovação (ANI), through the project Accelerat.AI candidature C644865762-00000008 funded by Recovery and Resilience Plan (RRP) and Next Generation EU European Funds.

References

[1] L.R. Horn, "A Natural History of Negation", University of Chicago Press, 1989.

[2] R.D. Huddleston and G.K. Pullum, "The Cambridge Grammar of the English Language", Cambridge University Press, 2002.

[3] R. Morante and E. Blanco, "Recent advances in processing negation", *Natural Language Engineering*, vol.27, no. 2, pp. 121-130, 2021.

[4] M.M. Hossain, D. Chinnappa and E. Blanco, "An Analysis of Negation in Natural Language Understanding Corpora", In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, vol. 2, pp. 716-723, May 2022.

[5] A. Mahany, H. Khaled, N.S. Elmitwally, N. Aljohani and S. Ghoniemy, "Negation and Speculation in NLP: A Survey, Corpora, Methods, and Applications", *Applied Sciences*, vol. 12. No. 10, p.5209, 2022.

[6] C. Dudschig, B. Kaup, M. Liu and J. Schwab, "The processing of negation and polarity: An overview", *Journal of Psycholinguistic Research*, vol. 50, no. 6, pp. 1199-1213, 2021.

[7] N. Adel, "Fuzzy natural language similarity measures through computing with words", *Doctorate of Philosophy*, Department of Computing and Maths, Manchester Metropolitan University (MMU), 2022.

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

[9] N. Adel, K. Crockett, D. Chandran and J.P. Carvalho, "Interpreting Human Responses in Dialogue Systems using Fuzzy Semantic Similarity Measures", *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp.1-8, 2020.

[10] E. Agirre, C. Banea, D. Cer, M. Diab, A. Gonzalez-Agirre, R. Mihalcea, G. Rigau and J. Wiebe, "Semeval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation", *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 497-511, 2016.

[11] H. Rubenstein and J. Goodenough, "Contextual Correlates of Synonymy", *Communications of the ACM*, vol. 8, pp. 627-633, 1965.

[12] D. Chandrasekaran and V. Mago, "Evolution of semantic similarity-A survey", *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1-37, 2021.

[13] N.A. Alnajran, "An integrated semantic-based framework for intelligent similarity measurement and clustering of microblogging posts", *Doctorate of Philosophy*, Department of Computing and Maths, Manchester Metropolitan University (MMU), 2019.

[14] N. Adel, K. Crockett, A. Crispin, J.P. Carvalho and D. Chandran, "Human Hedge Perception–and its Application in Fuzzy Semantic Similarity Measures", In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1-7, June 2019.

[15] N. Adel, K. Crockett, J.P. Carvalho. and V. Cross, "Fuzzy Influence in Fuzzy Semantic Similarity Measures" In *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1-7, July 2021.

[16] N. Adel, K. Crockett, D. Livesey and J.P. Carvalho, "An interval type-2 fuzzy ontological similarity measure", *IEEE Access*, vol. 10, pp.81506-81521, 2022.

[17] E. Sergeeva, H. Zhu, P. Prinsen and A. Tahmasebi, "Negation Scope Detection in Clinical Notes and Scientific Abstracts: A Feature-enriched LSTM-based Approach", *AMIA Summits on Translational Science Proceedings*, pp. 212-221, May 2019.

[18] O.S. Pabon, O. Montenegro, M. Torrente, A.R. Gonzalez, M. Provencio and E. Menasalvas, "Negation and uncertainty detection in clinical texts written in Spanish: a deep learning-based approach" *PeerJ Computer Science*, vol. 8, p.e 913, 2022.

[19] A. Pradhan, M.R. Senapati and P.K. Sahu, "Improving sentiment analysis with learning concepts from concept, patterns lexicons and negations" *Ain Shams Engineering Journal*, vol. 13, no. 2, p. 101559, 2022.

[20] U. Lal and P. Kamath, "Effective Negation Handling Approach for Sentiment Classification using synsets in the WordNet lexical database", *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, pp. 01-07, 2022.

[21] V. Novak, "Fuzzy logic in natural language processing" In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1-6, June 2017.

[22] A. Torrens-Urrutia, V. Novak, M.D. Jimenez-Lopez, "Describing Linguistic Vagueness of Evaluative Expressions Using Fuzzy Natural Logic and Linguistic Constraints" *Mathematics*, vol. 10, pp. 2760, 2022.

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

[24] E. Ferri, A. Kandel, G. Langholz, "Fuzzy Negation", In Zadeh, L.A., Kacprzyk, J. (eds) Computing with Words in Information/Intelligent Systems 1. Studies in Fuzziness and Soft Computing, vol. 33, Physica, Heidelberg, 1999.

[25] M. Sugeno, "Fuzzy measures and fuzzy integrals: a survey", In *Fuzzy Automata and Decision Processes*, pp. 89–102, North-Holland, Amsterdam, 1977.

[26] R.R. Yager, "On the measure of fuzziness and negation Part I: Membership in the Unit Interval" *International Journal of General Systems*, vol. 5, no. 4, pp. 221-229, 1979.

[27] [27] G.J. Klir and T.A. Folger, "Fuzzy sets, uncertainty, and information", Prentice-Hall, Inc., 1987.

[28] M. Higashi & G.J. Klir, "On measures of fuzziness and fuzzy complements", *International Journal of General Systems*, vol. 8, no. 3, pp. 169-180, 1982.

[29] Y. Li, D. McLean, Z.A. Bandar, J.D. O'Shea and 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.