OVERLAP SIMILARITY MEASURE BETWEEN INTUITIONISTIC FUZZY SETS AND ITS APPLICATIONS

ABSTRACT

A Similarity measure is a comparability measure that evaluates the degree of similarity between two items which has got applications in many areas like decision making, pattern recognition, word computing, clustering, etc. Even though numerous similarity measures between intuitionistic sets have been discussed in the literature, some fail to satisfy the similarity axioms whereas some do not consider the indeterminacy degree. To compensate for the disadvantage, in this paper, the overlap coefficient is used to form a new similarity measure that considers indeterminacy degree too. Specifically, the proposed method measures the degree of similarity between sets of different sizes. Its properties are discussed and finally few examples are given to delineate the practicality and adequacy of the proposed overlap similarity measure comparing it with the current comparability measures.

1. Introduction A Similarity measure or a similarity metric is a real-valued function used to measure the similarity between two items. Having a diverse variety of uses in the real world, similarity measure is applied in areas like image processing, pattern recognition, medical diagnosis, and decision making. Lotfi A. Zadeh independently proposed fuzzy sets in 1965 as an expansion of the standard notion of set[1]. The fuzzy set concept is appealing because it addresses uncertainty and ambiguity that the Cantorian set could not. It was extended to Intuitionistic fuzzy sets by Krassimir Atanassov in 1983 by including the degree of hesitation along with the membership and non-membership degree [2]. Szmidt and Kacprzyk in 1996 explained the application of IFS in decision making [3]. In 1990 Atanassov and Gargov applied the same in logic programming [4] and De et al. in medical diagnosis in 2001 [5]. One of the most important applications in pattern recognition came from Hung and Yang in 2004 [6].

Many researchers have explored the possibility of finding the similarity between intuitionistic fuzzy sets. One of the first applications of the similarity measure to pattern recognition problems was proposed by Li and Cheng. Concept of similarity measures between vague sets was proposed by Chen in 1995, 1997 [7]. In 1999 and 2001, two new similarity measures were introduced by Hong Kim and Fan Zhangyan respectively [8]. Jun Ye proposed the cosine similarity measure as well as the dice similarity measure between Intuitionistic fuzzy sets in 2011 and 2012 respectively [9][10]. Recently in 2018 Chao-Ming Hwang, Miin-Shen Yang, and Wen-Liang Hung introduced a new similarity measure based on the Jaccard index which gives application in clustering [11].

We discuss the Overlap similarity measure in this paper. It is defined as the ratio of the

AFSHAN

Research Scholar ,St.Thomas College, Palai,
afsh.libra@gmail.com

SHINY JOSE
Associate Professor in Mathematics,
St. George's College Aruvithura, Kerala,
shinyjosedavis@gmail.com

intersection of two sets to the size of the smaller set. Compared to other similarity measures Overlap Similarity measure stands out in the fact that it gives accurate results even when the sets are not of equal size. The rest of the paper is sectioned into the following parts. Section 2 discusses some of the basic definitions and prerequisites, section 3 deals with extending the Overlap Similarity measure in AIFS. Comparison of the proposed similarity measures with the existing measures is done in section 4 and its application in finding the similarity between neurodegenerative disorders and different languages is discussed in section 5. Finally, the results are summarised in the conclusion part in section 6.

2. Preliminaries

2.1 Definition

Let X be a given set. An Intuitionistic fuzzy set A in X is given by,

A = $\{(x, \mu_A(x), \nu_A(x)) | \, x \in X$ where $\mu_A, \nu_A: X \to [0,1]$, and $0 \le \mu_A(x) + \nu_A(x) \le 1$. $\mu_A(x)$ is the degree of membership of the element x in A and $\nu_A(x)$ is the degree of non membership of x in A. For each $x \in X$, $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ is called the degree of hesitation.

2.2 Definition

Let $\alpha, \beta \in$ IFS (χ). A mapping ς : IFS (χ) x IFS $(\chi) \longrightarrow [0,1]$ is said to be a degree of similarity between α and β if $\varsigma(\alpha,\beta)$ satisfy following properties

1. $0 \le \varsigma(\alpha,\beta) \le 1$
2. $\varsigma(\alpha,\beta) = 1$ if and only if $\alpha = \beta$
3. $\varsigma(\alpha,\beta) = \varsigma(\beta,\alpha)$
4. if $\alpha,\beta,\gamma \in$ IFS(χ) and $\alpha \subseteq \beta \subseteq \gamma$, then $\varsigma(\alpha,\gamma) \le \varsigma(\alpha,\beta), \varsigma(\alpha,\gamma) \le \varsigma(\beta,\gamma)$

2.3 Existing Similarity Measures

The following section lists the various similarity measures:

Concept of similarity measures between vague sets as proposed by Chen [7]

$$S_C(U,V) = 1 - \frac{\sum_{i=1}^{n} |S_U(x_i) - S_V(x_i)|}{2n} \quad (0.1)$$

In 1999 Hong and Kim defined S_H as follows [8]

$$S_H(U,V) = 1 - \frac{\sum_{i=1}^{n} |\mu_U(x_i) - \mu_V(x_i)| + |\nu_U(x_i) - \nu_V(x_i)|}{2n} \quad (0.2)$$

Fan and Zhangyan in 2001 proposed S_L as [9]

$$S_L(U,V) = 1 - \frac{\sum_{i=1}^{n} |S_U(x_i) - S_V(x_i)|}{4n} - \frac{\sum_{i=1}^{n} |\mu_U(x_i) - \mu_V(x_i)| + |\nu_U(x_i) - \nu_V(x_i)|}{4n} \quad (0.3)$$

S_O was proposed in 2002 by Yanhong et al [16]

$$S_O(U,V) = 1 - \sqrt{\frac{\sum_{i=1}^{n} (\mu_U(x_i) - \mu_V(x_i))^2 + (\nu_U(x_i) - \nu_V(x_i))^2}{2n}} \quad (0.4)$$

Cosine Similarity measure $C_{IFS}(U,V)$ was proposed by Jun Ye in 2011 as [10]

$$C_{IFS}(U,V) = \frac{1}{n} \sum_{i=1}^{n} \frac{\mu_U(x_i)\mu_V(x_i) + \nu_U(x_i)\nu_V(x_i)}{\sqrt{\mu_U(x_i)^2 + \nu_U(x_i)^2}\sqrt{\mu_V(x_i)^2 + \nu_V(x_i)^2}} \quad (0.5)$$

Dice Similarity measure $D_{IFS}(U,V)$ was proposed by Jun Ye in 2012 as [11]

$$D_{IFS}(U,V) = \frac{1}{n} \sum_{i=1}^{n} \frac{2(\mu_U(x_i)\mu_V(x_i) + \nu_U(x_i)\nu_V(x_i))}{\mu_U(x_i)^2 + \nu_U(x_i)^2 + \mu_V(x_i)^2 + \nu_V(x_i)^2} \quad (0.6)$$

A new Similarity measure based on the Jaccard index S_J was introduced in 2018 by Chao-Ming Hwang, Miin-Shen Yang, and Wen-Liang Hung [12].

$$S_J(U,V) = \frac{1}{3n}\sum_{i=1}^{n} (g(\mu_U(x_i), \mu_V(x_i)) + g(1-\nu_U(x_i), 1-\nu_V(x_i)) + g(\tfrac{1}{2}(1+\mu_U(x_i)-\nu_U(x_i), \tfrac{1}{2}(1+\mu_V(x_i)-\nu_V(x_i)) \quad (0.7)$$

$$where g(m,n) = \begin{cases} 1 & m=n=0 \\ \frac{m\times n}{m^2+n^2-m\times n} & m>0 \text{ and } n>0 \end{cases} \quad (0.8)$$

2.4 Definition

Overlap similarity measure is defined as the ratio of the intersection's size of two sets to the size of the smaller of the two sets.The overlap coefficient likewise called as Szymkiewicz-Simpson coefficient is a closeness measure that is identified with the Jaccard index.

$$OVL(C,D) = \frac{C.D}{min(\| C \|^2, \| D \|^2)} = \frac{\Sigma c_i d_i}{min(\Sigma c_i^2, \Sigma d_i^2)} \quad (0.9)$$

where C.D = $\Sigma c_i d_i$ and the L_2 norms of C and D is defined by $\|C\|_2 = \sqrt{\sum c^2}$ and $\|D\|_2 = \sqrt{\sum d^2}$

Overlap coefficient takes value in the range 0 to 1 and it is equal to1 when any one of the set is a subset of the other.

3. Overlap Similarity Measure This section deals with proposing a new similarity measure named overlap similarity measure and a weighted overlap similarity measure for AIFSs based on the idea of overlap coefficient between sets. Let C and D be two AIFSs within the discourse universe X. Let $\mu_C(\alpha_i)$, $\nu_C(\alpha_i)$, $\pi_C(\alpha_i)$ and $\mu_D(\alpha_i)$, $\nu_D(\alpha_i)$ $\pi_D(\alpha_i)$ be the membership degree, non membership degree and indeterminacy degree respectively of C and D respectively for $\alpha_i \in$ X

The overlap similarity measure between C and D is proposed as follows,

$$O_{AIFS}(C,D) = \frac{1}{n+1}\sum_{i=1}^{n} \frac{\mu_C(\alpha_i)\mu_D(\alpha_i)+\nu_C(\alpha_i)\nu_D(\alpha_i)+\pi_C(\alpha_i)\pi_D(\alpha_i)}{min(\mu_C(\alpha_i)^2+\nu_C(\alpha_i)^2+\pi_C(\alpha_i)^2, \mu_D(\alpha_i)^2+\nu_D(\alpha_i)^2+\pi_D(\alpha_i)^2)} \quad (0.10)$$

3.1 Preposition 1 The measure O_AIFS is a similarity measure between AIFS C and D.

Proof Inorder to prove that O_{AIFS} is a similarity measure, we need to show that it satisfies the following properties

1. 0≤ $O_{AIFS}(\alpha,\beta) \leq 1$
2. $O_{AIFS}(\alpha,\beta) = 1$ if and only if $\alpha = \beta$
3. $O_{AIFS}(\alpha,\beta) = O_{AIFS}(\beta,\alpha)$
4. if $\alpha,\beta,\gamma \in$ IFS(χ) and $\alpha \subseteq \beta \subseteq \gamma$, then $O_{AIFS}(\alpha,\gamma) \leq O_{AIFS}(\alpha,\beta)$, $O_{AIFS}(\alpha,\gamma) \leq O_{AIFS}(\beta,\gamma)$

1. Since $\mu_\alpha, \nu_\alpha, \mu_\beta, \nu_\beta \in$ [0,1] , (1) holds for every
2. $\mu_\alpha(x_i) = \mu_\beta(x_i)$, $\nu_\alpha(x_i) = \nu_\beta(x_i)$ whenever A = B .So $O_{AIFS} = 1$
3. proposition is true for obvious reasons
4.

$$O_{AIFS}(\alpha,\beta) = \frac{1}{n+1}\sum_{i=1}^{n} \frac{\mu_\alpha(x_i)\mu_\beta(x_i)+\nu_\alpha(x_i)\nu_\beta(x_i)+\pi_\alpha(x_i)\pi_\beta(x_i)}{min(\mu_\alpha(x_i)^2+\nu_\alpha(x_i)^2+\pi_\alpha(x_i)^2, \mu_\beta(x_i)^2+\nu_\beta(x_i)^2+\pi_\beta(x_i)^2)}$$

$$O_{AIFS}(\alpha,\gamma) = \frac{1}{n+1}\sum_{i=1}^{n} \frac{\mu_\alpha(x_i)\mu_\gamma(x_i)+\nu_\alpha(x_i)\nu_\gamma(x_i)+\pi_\alpha(x_i)\pi_\gamma(x_i)}{min(\mu_\alpha(x_i)^2+\nu_\alpha(x_i)^2+\pi_\alpha(x_i)^2, \mu_\gamma(x_i)^2+\nu_\gamma(x_i)^2+\pi_\gamma(x_i)^2)}$$

$$O_{AIFS}(\beta,\gamma) = \frac{1}{n+1}\sum_{i=1}^{n} \frac{\mu_\beta(x_i)\mu_\gamma(x_i)+\nu_\beta(x_i)\nu_\gamma(x_i)+\pi_\beta(x_i)\pi_\gamma(x_i)}{min(\mu_\beta(x_i)^2+\nu_\beta(x_i)^2+\pi_\beta(x_i)^2, \mu_\gamma(x_i)^2+\nu_\gamma(x_i)^2+\pi_\gamma(x_i)^2)}$$

When $\alpha \subseteq \beta \subseteq \gamma$, then $\mu(\alpha) \le \mu(\gamma)$ and $\nu(\alpha) \ge \nu(\gamma)$

Also $\mu(\beta) \le \mu(\gamma)$ and $\nu(\beta) \ge \nu(\gamma)$, $\mu(\alpha) \le \mu(\gamma)$ and $\nu(\alpha) \ge \nu(\gamma)$

So $O_{AIFS}(\alpha,\gamma) \le O_{AIFS}(\alpha,\beta)$ and $O_{AIFS}(\alpha,\gamma) \le O_{AIFS}(\beta,\gamma)$

If $\omega_i \in [0,1]$ $i = 1,2,\ldots\ldots$ n and $\Sigma\omega_i = 1$, a weighted Overlap similarity measure between C and D is proposed as follows,

$$WO_{AIFS}(C,D) = \sum_{i=1}^{n} w_i \frac{\mu_C(\alpha_i)\mu_D(\alpha_i)+\nu_C(\alpha_i)\nu_D(\alpha_i)+\pi_C(\alpha_i)pi_D(\alpha_i)}{min(\mu_C(\alpha_i)^2+\nu_C(\alpha_i)^2+\pi_C(\alpha_i)^2, \mu_D(\alpha_i)^2+\nu_D(\alpha_i)^2+\pi_D(\alpha_i)^2)} \quad (0.11)$$

When $w_i = \frac{1}{m+1}$ $i = 1,2\ldots\ldots$ n, then WO_{AIFS}(C,D) = O_{AIFS}(C,D)

And so the weighted Overlap similarity measure between AIFSs C and D also satisfies the following properties:

1. 0≤ $WO_{AIFS}(\alpha,\beta) \le 1$
2. $WO_{AIFS}(\alpha,\beta) = 1$ if and only if $\alpha = \beta$
3. $WO_{AIFS}(\alpha,\beta) = WO_{AIFS}(\beta,\alpha)$
4. if $\alpha,\beta,\gamma \in$ IFS(χ) and $\alpha \subseteq \beta \subseteq \gamma$, then $WO_{AIFS}(\alpha,\gamma) \le WO_{AIFS}(\alpha,\beta)$ $WO_{AIFS}(\alpha,\gamma) \le WO_{AIFS}(\beta,\gamma)$

The above properties can be proved similar to the previous proof.

4. Comparison of Similarity measures with the Overlap Similarity measure

Let A and B be two AIFS in X ={x_1, x_2, x_3, x_4}.

A={(x_1,0.2,0.1), (x_2,0.5,0.4), (x_3,0.3,0.2), (x_4,0.5,0.2)}

B={(x_1,0.3,0.3), (x_2,0.3,0.4), (x_3,0,0), (x_4,0.4,0.2)}

Calculation for various similarity measures are as follows:

S_C(A,B)=0.94
S_H((A,B)=0.86
S_L(A,B)=0.9
S_O(A,B)=0.83
C_{IFS}(A,B)=0.72
D_{IFS} (A,B)=0.675
S_J (A,B)=0.78

Overlap Similarity measure O(A,B) = 0.909

From the results, it can be interpreted that O and S_C have stronger similarity measures.

5. Applications

This section deals with the application of Overlap Similarity measure of AIFS in finding the similarity between four different Neurodegenerative disorders and similarity between different languages.

5.1 Example 1

Diseases which affect ones nervous system may be classified as Neurodegenerative disorders. Degenerative nerve conditions can be dangerous or even fatal. Depending on the kind many of them are incurable. Treatments may aid in alleviating symptoms, reducing discomfort, and enhancing mobility. Some of them are Alzheimer's disease, Huntington's disease, Parkinson's disease, Multiple system atrophy etc. Most of these conditions have common symptoms like memory loss, changes in mood, trouble with movement, difficulty in speaking and writing etc. Since these illness are fatal, it is adviced to take doctor's help in the early stages itself. In this paper, we consider the four most common Neurodegenerative disorders D = { Alzheimer's disease, Vascular Dementia, Parkinson's disease and Lewy Body Dementia } along with their common symptoms S = { memory loss, poor judgement, trouble interpreting visual information, and muffled speech }. Using the Overlap similarity measure ,the degree of similarity between each of these four conditions are obtained. Having an idea about how similar these diseases are will help to track the early symptoms. The Intuitionistic fuzzy relation between D and S is given in Table 1 and Table 2 gives the similarity between each of the diseases using various similarity measures.

Table 1: Symptoms characteristic for the diseases.

	Alzheimer's(D_1)	Vascular Dementia(D_2)	Parkinson's(D_3)	Lewy Body Dementia(D_4)
Memory loss	(0.8, 0.1)	(0.1,0.6)	(0.6,0.2)	(0.3,0.4)
Poor judgement	(0.6,0.3)	(0.2,0.4)	(0.8,0.2)	(0.7,0.1)
Trouble interpreting visual information	(0.5,0.3)	(0.3,0.6)	(0.7,0.2)	(0.6,0.2)
Muffled speech	(0.6,0.3)	(0.7,0.1)	(0.8,0.2)	(0.2,0.4)

Table 2: Similarity measures between each of the four diseases.

Similarity measure	$\varsigma(D_1, D_2)$	$\varsigma(D_1, D_3)$	$\varsigma(D_1, D_4)$	$\varsigma(D_2, D_3)$	$\varsigma(D_2, D_4)$	$\varsigma(D_3, D_4)$
S_C	0.69	0.85	0.77	0.69	0.66	0.81
S_H	0.69	0.85	0.77	0.69	0.68	0.82
S_L	0.68	0.85	0.79	0.68	0.67	0.83
S_O	0.63	0.84	0.73	0.62	0.64	0.76
D_{IFS}	0.35	0.56	0.41	0.38	0.31	0.31
C_{IFS}	0.58	0.91	0.85	0.79	0.68	0.86
S_J	0.64	0.78	0.72	0.69	0.65	0.74
O_{AIFS}	0.6	0.92	0.82	0.64	0.67	0.89

From Table 2 we can conclude that Alzheimer's disease and Parkinson's disease are more similar. It can also be clearly noted that the existing similarity measures also gives the same result though the proposed measure has a better degree of similarity since it takes into account the hesitancy degree.

5.2 Example 2

Researchers have been studying the commonalities across human languages for years. Lexical similarity is a linguistic term that refers to the degree to which two languages' word sets are similar. For a person who wishes to learn a new language, this similarity quotient will be of great help as it guides him in selecting the most easiest language depending on his mother tongue. Lets choose four South Indian languages namely Malayalam, Tamil, Telugu and Kannada. Using Overlap Similarity measure we show which of these languages is similar to Sanskrit so that a person wishing to learn Sanskrit has an idea about it. There have been four basic characteristics of language that have been investigated:- phonology, syn-tax, semantics, and pragmatics. Phonology is the study of sound patterns, Syn-tax, the structure of statements, Semantics, the meaning of words and Pragmatics is the study of how words are employed in everyday situations.

Consider a set of languages L = {L_1 (Malayalam), L_2 (Tamil), L_3 (Telugu), L_4 (Kannada)}, and a set of components of language R = { R_1 (Phonology), R_2 (Syn-tax), R_3 (Semantics), R_4 (Pragmatics)}. Suppose one wishes to learn a new language or choose his second language. Let Sanskrit be represented by the followig AIFS.

Q(Sanskrit)=
$\{(R_1, 0.1,0.6),\ (R_2, 0.2,0.8), (R_3, 0.6,0.1), (R_4, 0.8,0.1)\}$,

Each language $\mathrm{L_i}(i = 1,2,3,4)$can also be represented as AIFSs with respect to all components as follows:

$$L_1(Malayalam) = \left\{\begin{matrix}(R_1, 0.7,0.2), (R_2, 0.8,0.1), (R_3, 0.3,0.3), \\ (R_4, 0.5,0.2)\end{matrix}\right\},$$

$$L_2(Tamil) = \left\{\begin{matrix}(R_1, 0.4,0), (R_2, 0.2,0.8), (R_3, 0.2,0.7), \\ (R_4, 0.4,0.2)\end{matrix}\right\},$$

$$L_3(Telugu) = \left\{\begin{matrix}(R_1, 0.1,0.7), (R_2, 0.6,0.1), (R_3, 0.8,0), \\ (R_4, 0.6,0.4)\end{matrix}\right\},$$

$$L_4(Kannada) = \left\{\begin{matrix}(R_1, 0.4,0.3), (R_2, 0.5,0.4), (R_3, 0.6,0.1), \\ (R_4, 0.3,0)\end{matrix}\right\},$$

Degree of similarity between Sanskrit and the other four languages are calculated using Eq(0.10) and the following results are obtained.

$$O_{(L_{1,Q})} = 0.59,\quad O_{(L_{1,Q})} = 0.59,$$
$$O_{(L_3,Q)} = 0.732 \quad O_{(L_4,Q)} = 0.686.$$

Results show that Sanskrit and Telugu share maximum similarities between them and so a person familiar with the Telugu language will find it easier to learn Sanskrit.

6. Conclusion

One of the drawbacks found in the current similarity metrics was that the degree of hesitancy was not included along with the level of membership and non-membership. Overlap Similarity measure covers up that drawback as well as it proposes a measure between sets with different sizes. The suggested similarity metric when compared to the current measurements showed that the Overlap similarity measure was logical. Furthermore, applications illustrating similarities between diseases and languages are discussed. The extension of this work would be to find the deficiency in this method, rectify it, and then apply it in different fields.

References

1. L.A Zadeh, Fuzzy Sets, Information and Control , 8(1965) 338-353.

2. K.T Atanassov, Intuitionistic Fuzzy Sets, Elsevier North-Holland, Inc, USA, 20(1986).

3. Eulalia Szmidt, Janusz Kacprzyk, Remarks on some applications of intuitionistic fuzzy sets in decision making, Notes on Intuitionistic Fuzzy Sets , 2 (1996) 22-31.

4. Atanassov K, Gargov G, Interval valued intuitionistic fuzzy sets, Fuzzy Sets and Systems, 31(1989) 343–349.

5. De, Supriya Kumar, RanjitBiswas and Akhil Ranjan Roy, An Application of intutionistic fuzzy sets in medical diagnosis, Fuzzy Sets and Systems, 117(2001) 209-213.

6. Hung, Wen-Liang and Yang, Miin-Shen, Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance, Pattern recognition letters, Elsevier, 25(2004)1603--1611.

7. Chen, Similarity measures between vague sets and between elements, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 27 (1997) 153-158.

8. Dug Hun Hong, Chul Kim, A Note on Similarity Measures between Vague Sets and between Elements, Elsevier Science Inc., 115 (1999).

9. Fan, L., Zhangyan, Similarity measures between vague sets, J. Software, 12 (6)(2001), 922-927 (in Chinese).

10. Jun Ye, Cosine Similarity measures for Intuitionistic fuzzy sets and their applications, Mathematical and Computer Modelling, 53 (2011) 91-97.

11. Jun Ye, Multicriteria decision-making method using the Dice similarity measure based on the reduct intuitionistic fuzzy sets of interval-valued intuitionistic fuzzy sets, Applied Mathematical Modelling, 36(2012) 4466-4472.

12. Chao-Ming Hwang and Miin-Shen Yang, W. L. Hung, New similarity measures of intuitionistic fuzzy sets based on the Jaccard index with its application to clustering, International Journal of Intelligent Systems, 33 (2018) 1672-1688.

13. Donghai Liu, Xiao-hong Chen, Dan Peng, Cosine Similarity Measure between Hybrid Intuitionistic Fuzzy Sets and Its Application in Medical Diagnosis, Computational and Mathematical Methods in Medicine, (2018).

14. Dutta, Palash, An Advanced Dice Similarity Measure of Fuzzy Numbers and Its Application in Multicriteria Decision Making, 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), (2019) 1-6.

15. Li Dengfeng, Cheng Chuntian, New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions, Pattern Recognition Letters, 23 (2002) 221-225.

16. Li, Yanhong and Olson, David L and Qin, Zheng, Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis, Pattern Recognition Letters, Elsevier, 28 (2007) 278-285.