ASM Sc. J., 13, 2020
https://doi.org/10.32802/asmscj.2020.sm26(5.10)

Distance Measures of Intuitionistic Fuzzy Sets and its Application to Cancer Diagnosis

Sook Wern Chan[1] and Lazim Abdullah[2*]

[1]*School of Informatics and Applied Mathematics Universiti Malaysia Terengganu 21030 Kuala Nerus Malaysia*

[2]*School of Informatics and Applied Mathematics Universiti Malaysia Terengganu 21030 Kuala Nerus Malaysia*

Based on the concept of distance between two objects and the information carried by the membership degrees, non-membership degree and hesitancy degree of intuitionistic fuzzy sets (IFSs), this paper proposes the Euclidean distance measures between IFSs in a case of cancer diagnosis. For this purpose, a seven-step computational procedure is proposed for establishing Euclidean distance measures between two fuzzy relations. Three medical experts were requested to offer linguistic assessment of the symptoms with respect to cancer types. Assessment was also made to collect linguistic data of patients with respect to the symptoms. The IFSs linguistic terms are used as a toll in data collection. We also provide the comparative analysis between Euclidean distance measures and Hamming distance measures of IFSs. The two distance measures show a consistent result in suggesting a proper diagnosis for each patient. Further research could be undertaken to validate the results using different distance measures with the same datasets.

Keywords: Intuitionistic fuzzy set, Euclidean distance, Fuzzy relations, Hamming distance, Cancer diagnosis

I. INTRODUCTION

Distance is typically known as a measure of difference between two objects, points or sets. The most commonly used distance in basic mathematics is Euclidean distance. Mathematically, Euclidean distance is a measure of distance between two points in Euclidean space. Euclidean distance in basic mathematics is calculated by finding the square root of differences between the coordinates of two objects (Bandyopadhyay & Saha,2013). The most noticeable mathematics operations in Euclidean distance are square roots and square of deviations for its corresponding points. Apart from Euclidean distance, Hamming distance is another popular distance measure in two dimensions space. In the context of functions, Galatenko and Galatenko (2011) define Hamming distance as the number of coordinates in which corresponding vectors of values differ. The most noticeable mathematics operation that can be observed from the Hamming distance equation is the absolute values of deviations between two points. Differently from Euclidean distance, it has no square roots and also no squares of deviations. From mathematics operations perspective, it can be seen that the two distance measures used different mathematical operations. Therefore, it is hypothesized that distance measures obtained from Euclidean distance and Hamming distance are not consistent despite measuring the same objects or sets.

Distance measure is not only limited to a measure between two points, functions or objects, but also can be extended to sets. One of the sets that germane to uncertainty and fuzziness is intuitionistic fuzzy sets

*Corresponding author's e-mail:lazim_m@umt.edu.my

(IFSs). The IFS was proposed by Atanassov (1999) as a generalization to fuzzy sets. The IFS theory is based on extension of corresponding definitions of fuzzy sets objects and definition of new object and their properties (Atanassov,1999). Definition of distance between IFSs were fundamentally discussed by Szmidt, and Kacprzyk (2000). The use of distance between IFSs was further echoed by (Izadikhah,2009). Since then, distances of IFSs have attracted a handful of research in this field and have been used in many real-life applications. Hatzimichailidis et al., (2012) proposed a distance measure where information matrix norms with fuzzy implications can be applied to measure the distance between the IFSs. The proposed distance measure was applied to several pattern recognition problems. Ke et al., (2018) proposed a new distance measure between IFSs to overcome the drawbacks of some existing distances and applied it in decision making. They addressed the definition of an effective distance measure with concise form and specific meaning for Atanassov's IFS. A new distance measure for IFSs was defined based on a distance measure of interval values and the transformation from IFS to interval valued fuzzy sets. The distance measure was later combined with intuitionistic fuzzy TOPSIS to handle MCDM problems. Very recently, Liao et al., (2018) introduced distance measures between intuitionistic multiplicative sets to overcome the problems with existing distance measures. However, they omit the third membership for the simplicity of presentation. Yang and Chiclana (2012), investigated the consistency of two-dimensional distances and three-dimensional distances of IFSs. They advocated that three-dimensional distance functions are not required since the two-dimensional distance functions give a simple and clear expression of the distance between two IFSs.

Most of the distance measures of IFSs have been applied to pattern recognitions. Some researchers also apply the IFSs based distance measures to medical sciences. Davarzani & Khoreh (2013), for example, studied on four new distance measures of IFSs and its applications in pattern recognition. They applied the concept of distance measures to medical diagnosis progress of bacillus colonies identification. In another research, distance measures of IFSs also have been applied to medical diagnosis. Djatna et al., (2018), applied IFSs-Hamming distance based on decision tree to diagnosis the different types of stroke disease. Luo, and Zhao (2018) proposed a new distance measure between IFS sand the algorithm based on these distance measures was developed for the use in medical diagnosis problems.

From a computational point of view, one of the notable discoveries from the work of Szmidt, and Kacprzyk (2000) is the way the distance measure of IFS was calculated. They suggested that all three memberships that characterizing IFSs must be considered in calculating distance measures. Considering the information contained by three membership degrees of IFSs as vector representations in the vector space, the objective of this paper is to propose Euclidean distance measures and Hamming distance measures of IFSs for the case of cancer diagnosis. According to National Cancer Society Malaysia (National Cancer Society Malaysia, 2015), there are five cancers that affecting both men and women in Malaysia. The types of cancer are namely breast cancer, colorectal cancer, lung cancer, cervical cancer, and nasopharyngeal cancer. There are many different symptoms of cancer. However, most of these types of cancer shared common symptoms such as unexplained bleeding, unexplained weight loss, a lump or swelling, and unexplained pain. Detailed application of distance measures in cancer diagnosis are further explained in section III.

II. PRELIMINARIES

Definitions of IFSs and two distance measures of IFSs are defined in this section. These definitions are required prior to applying distance measures of IFSs for the case of cancer diagnosis.

Definition 2.1 Intuitionistic Fuzzy Sets (Atanassov,1999).

An IFS A in X is defined as:

$$A = \{ \langle \ x, \mu_A(x), v_A(x) \ \rangle \mid x \in X \ \},$$

where

$$\mu_A : X \to [0,1]$$

and

$$\nu_A : X \rightarrow [0,1]$$

with the condition

$$0 \le \mu_A(x) + \nu_A(x) \le 1 \; \forall x \in X$$

The functions $\mu_A(x)$ and $\nu_A(x)$ denote the degree of membership and the degree of non-membership of the element $x \in X$, respectively. Obviously, every ordinary fuzzy set can be described as $\{ \langle x, \mu_A(x), 1-\mu_A(x) \rangle \mid x \in X \}$.

Definition 2.2 Intuitionistic Fuzzy Sets with Hesitation degree (Atanassov,1999).

For an intuitionistic fuzzy set with hesitation degree A',

$$A' = \{ \langle x, \mu_A(x), \nu_A(x), \pi_A(x) \rangle \mid x \in X \},$$

a hesitation margins

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x),$$

which is the intuitionistic fuzzy index of the element $x \in A$, symbolises a measure of the degree of non-determinacy of the element $x \in X$. It expresses the lack of knowledge on whether x belongs to A or not. Apparently, $0 \le \pi_A(x) \le 1$, for every $x \in X$. For every fuzzy set A',

$$\pi_{A'}(x) = 1 - \mu_{A'}(x) - [1 - \mu_{A'}(x)] = 0$$

where $x \in X$.

Definition 2.3: Euclidean Distance of IFSs (Szmidt & Kacprzyk, 2000).

Consider two IFSs A and B in $X = \{x_1, x_2, \ldots, x_n\}$. The Euclidean distance for IFSs is defined as

$$d_{EIFS}(A,B) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}(\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2}$$

Definition 2.4: Hamming Distance of IFSs (Szmidt & Kacprzyk, 2000).

For the two IFSs A and B in $X = \{x_1, x_2, \ldots, x_n\}$, the Hamming distance is

$$d_{HIFS}(A,B) = \frac{1}{2n}\sum_{i=1}^{n}|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|$$

Property of distance (Szmidt & Kacprzyk, 2000)

Any set *X* whose elements are called points is said to be a metric space if for any two points *A, B*∈*X*, there is an associated real number *d(A, B)* called distance if the following properties hold true:

i. *d(A, B)* ≥ 0, (non-negativity)

ii. d(A, B) = 0 ⇔ *A = B* (definiteness)

iii. d(*A, B*) = *d*(*B, A*) (symmetry)

iv. d(*A, C*) ≤ *d*(*A, B*) + *d*(*B, C*), for any *B*∈*X* (triangle inequality).

Any functions with all these properties are called distance functions, or distance measures.

Based on these properties, the word 'distance' and the words 'distance measure' are used interchangeably in this paper. These preliminaries are used prevalently in the case of identification of cancer types using the concepts of distance measures.

III. PROPOSED WORK

This section describes the application of distance measure to the case of cancer diagnosis. The identification of cancer types is made by examining the distances between symptoms and patients, and also distances between patients and cancer types. Linguistic variables of IFSs employed by experts, the proposed algorithm that purposely used for the case, and also the results are explained in the following sub-sections.

A. Experts and linguistic variables

In this research, a committee of three experts was requested to offer qualitative evaluation of cancer diagnosis using the linguistic variables adopted from Chang and Chen (1994). The experts comprise three doctors attached to Malaysia Ministry of Health in Ipoh Malaysia. The main points retrieved from personal communication were their linguistic evaluations pertaining to the cancer symptoms that possibly could be associated with different types of cancer.

B. Proposed Euclidean distance measures of IFSs

The computation steps that apply the Euclidean distance measures of IFSs in cancer diagnosis are proposed as follows.

Input

Identify a set of patients $P = \{P_1, \ldots, P_m\}$

Construct a set of symptoms $S = \{S_1, \ldots, S_n\}$.

Construct a set of cancer types $D = \{D_1, \ldots, D_q\}$.

Process

Step 1: Create an intuitionistic fuzzy relation A from the set of patients to the set of symptoms. For each patient, $P_i, i = 1, \ldots m$, a set of symptoms $S_j, j = 1, \ldots n$ is specified. The fuzzy relation A is defined as

$$A = \{\langle (p,s), \mu_A(p,s), v_A(p,s), \pi_A(p,s) | (p,s) \in P \times S \rangle\}$$

where $\mu_A(p,s)$ indicates the degree of the symptom s appears in patient p, $v_A(p,s)$ indicates the degree of the symptom s does not appear in patient p, and $\pi_A(p,s)$ symbolises the degree of uncertainty of the presence of the symptom s in patient p. It shows the degree of relationship between the patients and the symptoms.

Step2: Construct an intuitionistic fuzzy relation B from the set of symptoms to the set of cancer types. Consider a set of cancer types $D = \{D_1, \ldots, D_q\}$. For each cancer type $D_k, k = 1, \ldots q$, a set of symptoms $S_j, j = 1, \ldots n$ is specified. The fuzzy relation B is defined as

$$B = \{\langle (s,d), \mu_B(s,d), v_B(s,d), \pi_B(s,d) | (s,d) \in S \times D \rangle\}$$

where $\mu_B(s,d)$ indicates the degree of the symptom s confirms the presence of the cancer type d, $v_B(s,d)$ indicates the degree of the symptom s confirms the non-existence of the cancer type d, and $\pi_A(p,s)$ denotes the degree of uncertainty of the symptom s confirms the presence of the cancer type d. It shows the degree of relationship between the symptoms and the diseases, which is also known as conformability degree.

Step 3: Calculate the Euclidean distance measures for all symptoms of the i th patient from the k th disease using the Definition 2.3.

Output

Identify the diagnosis based on the distances between two intuitionistic fuzzy relations.

The algorithm is executed in the case of patients that seeking an appropriate diagnosis. The results are presented in the next sub-section.

C. Results

A one to one mapping fuzzy relations of the set of patients to the set of cancer types are constructed. Using these two fuzzy relations, Euclidean distance measures can be computed. Detailed computations are made according to the following steps.

Input

A set of patients,

P= {Fati, Lam, Kuz, Shei, Hal}

A set of symptoms,

S= {Unintended Weight Loss, Unusual Bleeding, Swelling or Lump, Shortness of Breath, Persistent Cough, Fatigue, Diarrhea, Swallowing Problem, Loss of Appetite}

Process

Step 1: An intuitionistic fuzzy relation A from the set of patients to the set of symptoms is computed.

Step 2: A set of cancer types *D = {Lung Cancer, Breast Cancer, Colorectal Cancer, Nasopharyngeal Cancer, Cervical Cancer}*

Step 3: An intuitionistic fuzzy relation B from the set of symptoms to the set of cancer types is computed.

Output

Distance measure of IFSs for all diagnosis of the i-patient from k-th disease are attained using the equation in Definition 2.3. The results of the Euclidean distance measures of IFSs are presented in Table 1.

Table 1. Euclidean Distance Measures between Patients and Cancer Types

	Lung Cancer	Breast Cancer	Colorectal Cancer	Nasopharyngeal Cancer	Cervical Cancer
Fati	0.3844	0.3667	0.1826	0.2769	0.1414
Lam	0.4295	0.4123	0.0816	0.3958	0.1563
Kuz	0.2687	0.2867	0.4269	0.1453	0.3873
Shei	0.1915	0.1453	0.4177	0.2789	0.4082
Hal	0.0943	0.1633	0.4256	0.3018	0.4041

The shortest distance points out a proper diagnosis for each patient. From the Table 1, it can be seen that Fati suffers from cervical cancer, Lam suffers from colorectal cancer, Kuz suffers from nasopharyngeal cancer, Shei suffers from breast cancer and Hal suffers from lung cancer.

The similar algorithm is iterated for Hamming distance measures of IFSs for the same sets of patients, symptoms and cancer types. Step 6 in the algorithm is now replaced by Hamming distance equation in Definition 2.4.

The distance measures obtained from the two difference equations are compared. Figure 1 shows the distance measures for each patient under the Euclidean distance measures and Hamming distance measures respectively.

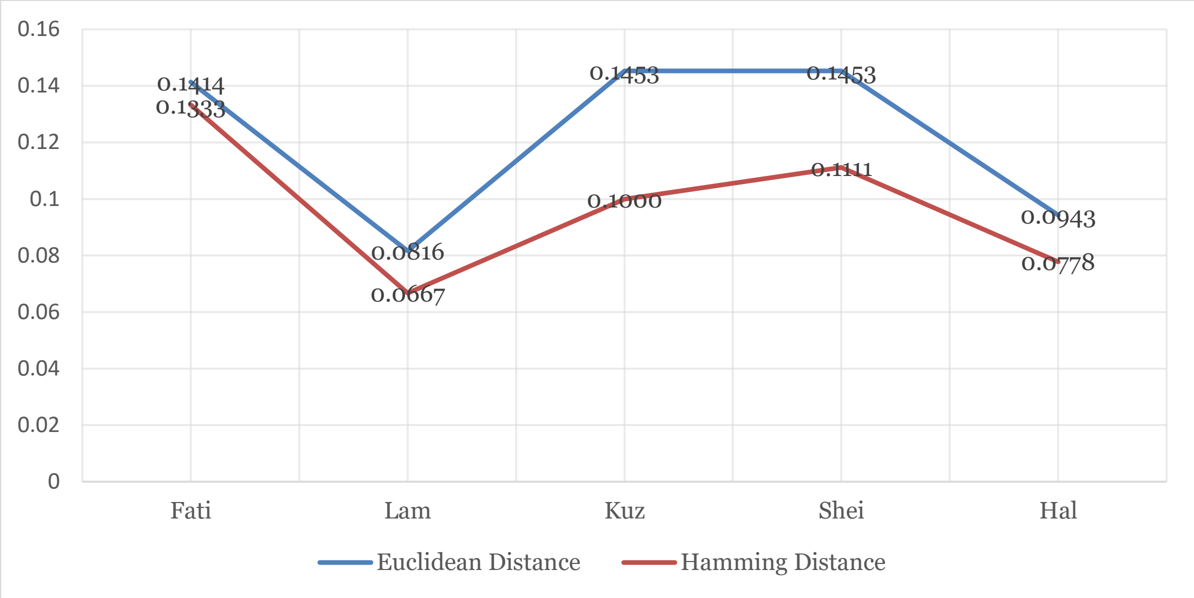

Figure 1 Distance Measures of Patients

The graph indicates that the identification of cancer diagnosis for each patient is consistent despite differences in values of distance measures. It can be concluded that the Eulidean distance measures and Hamming distances measures are consistent in suggesting the cancer diagnosis for patients.

IV. CONCLUSIONS

Distance is typically used to find dissimilarity between two objects or sets. It is defined in various equations depending on dimensions or spaces where distances are measured. In this paper, we have proposed Euclidean distance measures between two intuitionistic fuzzy relations that representing one to one mapping of patients and symptoms, and also patients and cancer types. The same data sets were also used to find Hamming distance measures. The two distance measures were compared, and the results show the consistency in suggesting cancer diagnosis.

The result somehow disapproved the belief that Euclidean distance measures and Hamming distance measures are dissimilar in terms of providing cancer diagnosis. However, the values of distance measures between Euclidean distance measures and Hamming distance measures are not identical.

This supports the notion that these two distance measures work under different mathematics operations. Future research can be undertaken to deeply explore the mathematical operations behind these two distance measures. It is also noticed that the gaps between these two distance measures are not consistent, and therefore, gap analysis and error analysis could be explored in future research.

V. REFERENCES

Atanassov, K. T. (1999). Intuitionistic fuzzy sets: theory and application. Heidelberg: Physica-Verlag.

Bandyopadhyay, S. and Saha, S. (2013). Unsupervised classification: similarity measures, classical and metaheuristic approaches, and applications. Berlin: Springer.

Chang, P. L., and Chen, Y. C. (1994). A fuzzy multi-criteria decision-making method for technology transfer strategy selection in biotechnology. Fuzzy Sets and Systems, 63(2), 131- 139.

Davarzani, H., and Khorheh, M. A. (2013). A novel application of intuitionistic fuzzy sets theory in medical science: Bacillus colonies Cognition. Artificial Intelligence Research 2(2), 1-16.

Djatna, T., Hardhienata, M.K.D, and Nur Masruriyah A.F. (2018). An intuitionistic fuzzy diagnosis analytics for stroke disease, Journal of Big Data, 5, 35, https://doi.org/10.1186/s40537-018-0142-7.

Galatenko, A.V. and Galatenko, V.V. (2011) Journal of Mathematical Sciences 172, 650. https://doi.org/10.1007/s10958-011-0210-4.

Hatzimichailidis, A.G., Papakostas, G.A., and Kaburlasos, V.G. (2012). A novel distance measure of intuitionistic fuzzy sets and its application to pattern recognition applications, International Journal of Intelligent Systems. 27 (4), 396–409.

Izadikhah, M. (2009). Using the Hamming distance to extend TOPSIS in a fuzzy environment. Journal of Computational and Applied Mathematics 231(1), 200-207.

Ke, D., Song, Y., and Quan, W. (2018). New Distance Measure for Atanassov's Intuitionistic Fuzzy Sets and Its Application in Decision Making. Symmetry 2018, 10, 429.

Liao, H., Zhang, C., and Luo, L. (2018). A multiple attribute group decision making method based on two novel intuitionistic multiplicative distance measures. Information Sciences, 467, 766-783.

Luo, M. and Zhao, R. (2018). A distance measure between intuitionistic fuzzy sets and its application in medical diagnosis, Artificial Intelligence in Medicine, 89, 34-39.

National Cancer Society Malaysia. (2015). Types of cancer. http://www.cancer.org.my/quick-facts/types-cancer/. [17 October 2015].

Szmidt, E. and Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets. Fuzzy Sets and Systems 114(3), 505-518.

Yang, Y. and Chiclana, F. (2012). Consistency of 2D and 3D distances of intuitionistic fuzzy sets. Expert Systems with Applications 39(10), 8665-8670.