A Novel Similarity Measure and Its Application in Medical Diagnosis under Intuitionistic Fuzzy Settings

Lishi Zhang
School of Science, Dalian Ocean University, Dalian, China
zls@dlou.edu.cn

Keywords: Pattern recognition; Intuitionistic fuzzy soft set; Attribute; weights; Similarity measure; Group decision makings

Abstract. The purpose of this paper is to develop a novel approach for medical diagnosis reasoning, in which a novel similarity measure is defined and used in the application of medical pattern recognition; the information is in the form of intuitionistic fuzzy numbers. First, a new biparametric similarity measure is defined, then, the similarity between the disease and the patients are computed, and an illustrative example is provided to illustrate its reasonability and validness of the proposed approach.

Introduction

As a generalization of fuzzy sets [1], intuitionistic fuzzy sets [2-3] can act as a better model to handle uncertain and fuzzy information which is involved in the decision making process. In the medical diagnosis process, the related information is presented in the form of intuitionistic fuzzy sets which are tools for a proper presentation of medical information under imperfectly and imprecise condition. In specific, The data set is listed as follows: 1.A set of symptoms $S$ which are the factors which cause some kinds of disease. 2: A set of disease $D$ which a fixed group of patients may suffer. The information value is presented in the light of intuitionistic fuzzy set that makes it more suitable to express imperfect information. In some cases, information cannot be acquired directly due to lack of background knowledge, unavailability of measuring tools, under time pressure. In such a case, the use of a degree of membership and non-membership can be a sufficient representation approach. A lot of applications [4-7] have been carried out in various fields. The proposed method of medical diagnosis develops a new measure of similarity for intuitionistic fuzzy sets. The similarity measures between the set of symptoms for each patient and a set of symptoms are computed. By ranking the measure of similarity, we get the highest similarity measures which corresponds the diagnosis that the patient is most likely to suffer from.

The rest of the manuscript is organized as follows. In Section 2 briefly recalls some background knowledge about intuitionistic fuzzy sets. In Section 3 we propose a new measure of similarity for intuitionistic fuzzy sets. In Section 4 gives an illustrative example to testify our proposed methods and the comparison with the previous approach is made.

Preliminaries

In this section, we will briefly recall some basic concepts and notions; background knowledge of interval valued intuitionistic fuzzy sets is introduced.

Definition 2.1. [1] Let $X$ a set, an intuitionistic fuzzy set (IFS) $A$ in $X$ is defined as $A = \{(x, \mu_A(x), \nu_A(x)) \mid x \in X\}$

Where $\mu_A(x) \in [0,1]$ and $\nu_A(x) \in [0,1]$ are the degree of membership and nonmembership of an element $x \in U$ respectively, satisfying $0 \leq \mu_A(x) + \nu_A(x) \leq 1$, $\forall x \in X$. $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ is called the intuitionistic fuzzy index of $x \in U$. Each fuzzy set can be represented as the following intuitionistic fuzzy set, in paper [8], a new distance measure is defined based on the centroid points of transformed right-angled triangular fuzzy numbers

Definition 2.1 [8] Let $A, B$ be two IVIFVs in $X = \{x_1, x_2, \cdots, x_n\}$, the distance between $A$ and $B$ is
\[
D(A,B) = \sqrt[p]{\frac{1}{2n(1+t)^p} \sum_{i=1}^{n} [\left( t(\mu_A(x_i) - \mu_B(x_i)) + |V_A(x_i) - V_B(x_i)| \right)^p + \left( |t(\mu_A(x_i) - \mu_B(x_i))| + |V_A(x_i) - V_B(x_i)| \right)^p]}
\]

To utilize more information of the intuitionistic fuzzy sets, we improve the distance measure as follows, we add the items

\[
(|t(\pi_A(x_i) - \pi_B(x_i))| + |\theta_A(x_i) - \theta_B(x_i)|)^p;
\]

\[
(|t(\theta_A(x_i) - \theta_B(x_i))| + |\pi_A(x_i) - \pi_B(x_i)|)^p;
\]

\[
\pi_A(x_i) = 1 - \mu_A(x_i) - v_A(x_i), \pi_B(x_i) = 1 - \mu_B(x_i) - v_B(x_i)
\]

\[
\theta_A(x_i) = \mu_A(x_i) + v_A(x_i), \theta_B(x_i) = \mu_B(x_i) + v_B(x_i)
\]

For the convenience, the improved distance measure is defined as

\[
c_{ij} = (|t(\mu_A(x_i) - \mu_B(x_i)) + |V_A(x_i) - V_B(x_i)|)^p; c_{i2} = (|t(\mu_A(x_i) - \mu_B(x_i))| + |V_A(x_i) - V_B(x_i)|)^p
\]

\[
c_{i3} = (|t(\pi_A(x_i) - \pi_B(x_i))| + |\theta_A(x_i) - \theta_B(x_i)|)^p; c_{i4} = (|t(\pi_A(x_i) - \pi_B(x_i))| + |\theta_A(x_i) - \theta_B(x_i)|)^p
\]

**Definition 2.2** Let \( A, B \) be two IVIFVs in \( X = \{x_1, x_2, \ldots, x_n\} \), the distance between \( A \) and \( B \) is

\[
D(A,B) = \sqrt[p]{\frac{1}{2n(1+t)^p} \sum_{i=1}^{n} [c_{ij} + c_{i2} + c_{i3} + c_{i4}]}
\]

\[
S(A,B) = 1 - \sqrt[p]{\frac{1}{2n(1+t)^p} \sum_{i=1}^{n} [c_{ij} + c_{i2} + c_{i3} + c_{i4}]}
\]

Are the distance measure and similarity measure.

**The Proposed Algorithms**

The following assumptions or notations are used to represent the MADM problems with known weight information in intuitionistic fuzzy setting.

Let \( P = \{p_1, p_2, \ldots, p_m\} \) be a set of patients, and \( S = \{s_1, s_2, \ldots, s_k\} \) be a set of symptoms, and \( D = \{d_1, d_2, \ldots, d_n\} \) be a set of diseases.

\[
PS_{mk} = (a_{ij})_{m \times k}, a_{ij} = [\mu_{ij}(p_i), V_{ij}(p_i)], i = 1, 2, \ldots, m; j = 1, 2, \ldots, K
\]

Where \( a_{ij} \) denotes the degree value that patient \( p_i \) shows the symptom \( s_j \).

\[
DS_{mk} = (b_{ij})_{n \times k}, b_{ij} = [\mu_{ij}(d_i), V_{ij}(d_i)], i = 1, 2, \ldots, n; j = 1, 2, \ldots, K
\]

Where \( b_{ij} \) denotes the degree value that symptom \( s_j \) indicates the disease \( d_j \). By using the similarity measure in Definition 2.2, we get the similarity measure matrix

\[
SM_{mk} = (s_{ij})_{m \times k}, s_{ij} = S(p_i, d_j), i = 1, 2, \ldots, n; j = 1, 2, \ldots, K
\]

If

\[
s_{ij} = \max\{s_{ij} \mid j = 1, 2, \ldots, K\}
\]

Then patient \( p_i \) suffers from disease \( d_j \).
Case Study

Example 1 [9]: Let there be four patients Paul, Jadu, Kundu, and Rohit in hospital at Calcutta. Their symptoms are temperature, headache, stomach pain, cough, and chest-pain. Clearly, the set of patients $P = \{ \text{Paul, Jadu, Kundu, Rohit} \} = \{ p_1, p_2, p_3, p_4 \}$.

The set of symptoms $S = \{ \text{temperature, headache, stomach pain, cough, chest-pain} \} = \{ s_1, s_2, s_3, s_4, s_5 \}$.

The set of disease $D = \{ \text{Viral Fever, Malaria, Typhoid, Stomach problem, Heart problem} \} = \{ d_1, d_2, d_3, d_4, d_5 \}$.

The intuitionistic fuzzy soft sets of patients and symptoms are represented as $F(e_i) = \{ (h_1, [0.8, 0.1]), (h_2, [0, 0.8]), (h_3, [0.8, 0.1]), (h_4, [0.6, 0.01]) \}$.

The intuitionistic fuzzy soft sets of symptoms with respect to disease are represented as $G(d_j) = \{ (e_1, [0.4, 0]), (e_2, [0.3, 0.5]), (e_3, [0.1, 0.7]), (e_4, [0.4, 0.3]), (e_5, [0.1, 0.7]) \}$.

The matrix presentation is $X_{4 \times 5} = \begin{bmatrix}
0.8,0.1 & 0.6,0.1 & 0.2,0.8 & 0.6,0.1 & 0.1,0.6 \\
0.0,0.8 & 0.4,0.4 & 0.6,0.1 & 0.1,0.7 & 0.1,0.8 \\
0.8,0.1 & 0.8,0.1 & 0.0,0.6 & 0.2,0.7 & 0.0,0.5 \\
0.6,0.1 & 0.5,0.4 & 0.3,0.4 & 0.7,0.2 & 0.3,0.4
\end{bmatrix}^4_{4 \times 5}$.

The intuitionistic fuzzy soft sets with respect to disease are represented as $G(d_j) = \{ (e_1, [0.7, 0]), (e_2, [0.2, 0.6]), (e_3, [0, 0.9]), (e_4, [0.7, 0]), (e_5, [0.1, 0.8]) \}$.

The matrix presentation is $Y_{5 \times 5} = \begin{bmatrix}
0.4,0.1 & 0.7,0.0 & 0.3,0.3 & 0.1,0.7 & 0.8,0.1 \\
0.3,0.5 & 0.2,0.6 & 0.6,0.1 & 0.2,0.4 & 0.0,0.8 \\
0.1,0.7 & 0.0,0.9 & 0.2,0.7 & 0.0,0.8 & 0.2,0.8 \\
0.4,0.3 & 0.7,0.0 & 0.2,0.6 & 0.2,0.7 & 0.2,0.8 \\
0.1,0.7 & 0.1,0.8 & 0.1,0.9 & 0.2,0.7 & 0.8,0.1
\end{bmatrix}^5_{5 \times 5}$.

With the similarity measure matrix, we get that

$SM = \begin{bmatrix}
\text{Paul} & \text{Viral Fever} & \text{Malaria} & \text{Typhoid} & \text{Stomach} & \text{Heart problem} \\
0.9142 & 0.7748 & 0.9736 & 0.7494 & 0.6995 \\
0.6982 & 0.9000 & 0.7494 & 0.9736 & 0.6500 \\
0.9142 & 0.7249 & 0.9646 & 1.0000 & 0.6496 \\
0.9646 & 0.8747 & 0.8986 & 0.7500 & 0.7993
\end{bmatrix}$

From the above table, we see that Jadu, Kundu, Rohit suffer from the disease stomach, while Paul suffers from the disease Typhoid, Rohit suffers from the disease Viral Fever. Let us observe another example.

Example 2 [10]: Let us consider the following pattern recognition problem as discussed in [10]. Given five known patterns, which corresponds to five decision alternatives $R_1, R_2, R_3, R_4, R_5$ respectively. The patterns are denoted by the following IVIF.
\[ R_1 = \{ (h_1,[0.4,0.4]), (h_2,[0.4,0.4]), (h_3,[0.3,0.5]), (h_4,[0.5,0.3]) \} \]
\[ R_2 = \{ (h_1,[0.5,0.3]), (h_2,[0.6,0.3]), (h_3,[0.5,0.4]), (h_4,[0.4,0.2]) \} \]
\[ R_3 = \{ (h_1,[0.3,0.4]), (h_2,[0.1,0.6]), (h_3,[0.2,0.5]), (h_4,[0.2,0.6]) \} \]
\[ R_4 = \{ (h_1,[0.2,0.4]), (h_2,[0.4,0.4]), (h_3,[0.4,0.5]), (h_4,[0.5,0.2]) \} \]
\[ R_5 = \{ (h_1,[0.3,0.3]), (h_2,[0.7,0.2]), (h_3,[0.5,0.4]), (h_4,[0.6,0.2]) \} \]

Given an unknown sample
\[ r = \{ (h_1,[0.5,0.3]), (h_2,[0.7,0.2]), (h_3,[0.5,0.4]), (h_4,[0.6,0.2]) \} \]

Our aim is to classify pattern \( r \) to one of the decision alternatives \( R_1, R_2, R_3, R_4, \) and \( R_5 \). That is, we will find which one of the five decisions is mostly close to the unknown sample \( r \). The order of similaritis are listed as follows
\[ S(R_3, r) = 0.9473, S(R_4, r) = 0.8973, S(R_5, r) = 0.7432, S(R_1, r) = 0.6993, S(R_2, r) = 0.3726 \]
\[ S(R_5, r) > S(R_1, r) > S(R_2, r) > S(R_4, r) > S(R_3, r) \]

It means that the decision alternative \( R_5 \) is the optimal alternative. This result is in agreement with the one obtained in [10].

References