Journal of Nonlinear and Convex Analysis
Volume 26, Number 4, 2025, 865–881

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

A NOVEL INTUITIONISTIC FUZZY SOFT METRIC AND ITS APPLICATION TO CLUSTERING

NIHER RANJAN DAS, ASHUTOSH TIWARI, AND Q. M. DANISH LOHANI

Abstract. The intuitionistic fuzzy set (IFS) replicates the dataset of the clustering problem in a three-dimensional space which ultimately helps in the deeper analysis of the dataset. The intuitionistic fuzzy soft set (IFSS) involves an addition of the parameter within IFS to enhance its versatility over real-world problems of pattern recognition, machine learning, and decision-making. Here, a concept of bounded variation of IFSS is exploited for the proposal of a metric space. In the paper, we introduce an IFSS-based similarity measure with the help of the proposed metric space to perform the hierarchical clustering. We applied this approach to a software dataset and conducted a detailed comparative analysis of the clustering results using benchmark indexes.

1. Introduction

In many real-world scenarios, a variety of uncertainties is commonly encountered. The traditional mathematical approach proves insufficient for addressing these uncertainty challenges in the contemporary context. Consequently, fuzzy set theory and intuitionistic fuzzy set theory are treated as a logic system to address intricate problems involving multiple uncertainties. Atanassov [1] introduced the concept of IFSs as an extension of fuzzy sets (FSs) [33]. The intuitinistic fuzzy environment can encapsulate indecision while measuring membership degree and non-membership degree to help the decision-making. The application of intuitionistic fuzzy set theory is widespread across diverse fields, including decision making [8], machine learning [16, 29], pattern recognition [3], medical diagnosis [26], clustering [17], and various other application domains. The absence of appropriate parameterization tools in FS and IFS theories resulted a proposal of IFSS. The IFSS addresses the limitations of FS and IFs by delineating a more specialized category of subsets within the universe of discourse. Molodtsov [21] showed the adaptability of soft sets as a method for expressing and assessing problems involving uncertainty. Maji [18] integrated the concepts of SS and IFS, resulting in the creation of IFSS as a more dependable extension of IFS theory. Dinda and Samanta [9] introduced the concept of IFSS relations, exploring various algebraic features based on the work of Maji et al. Subsequently, Majumdar and Samanta [19] introduced diverse similarity measures based on SS, resembling the measure initially proposed by Chen [6]. They extended the concepts of FSS and IFSS to generalized fuzzy sets (GFSS) and generalized intuitionistic fuzzy soft sets (GIFSS) [20].

2020 *Mathematics Subject Classification.* 03B52, 03E72, 11J83.
Key words and phrases. Soft set, Intuitionistic fuzzy soft set, Similarity measure, Clustering algorithms.

Distance/similarity measures serve as efficient tools for evaluating the degree of similarity or dissimilarity between two sets. They find widespread application across various academic domains, including pattern recognition, machine learning, decision-making, signal detection, security verification systems, and image processing. Several similarity measures grounded in IFS have been proposed. Szmidt and Kacprzyk [25] presented diverse similarity measures for IFSs, utilizing metrics like Euclidean distance, Hamming distance, and their normalized counterparts. Hung and Yong [14] integrated the l_p metric and Hausdorff distance concepts to introduce a novel similarity measure for IFSs. Identifying limitations in existing IFS similarity measures, Liang and Shi [15] proposed a specific measure tailored for addressing pattern recognition problems. Chen [5] introduced various measures for assessing the similarity between vague sets and their elements. Hong and Kim [12] illustrated instances where Chen's similarity measures [5] are not always effective, suggesting modified measures in such scenarios. Cagman and Deli [4] developed a decision-making method for addressing medical diagnosis problems using IFSSs. More recently, Muthukumar and Krishnan [22] introduced a distinct type of similarity measure and a weighted similarity measure based on IFSSs, along with a discussion on their mathematical properties. Hung and Yang (2004) [13] established a link between similarity and distance measures, demonstrating their complementary nature. Similarly, Wang and Xin [30] scrutinized the equality between similarity and distance measures, proposing alternative distance measures by presenting counterexamples that rendered the distance measures in [25] impractical. Moreover, numerous distance measures for IFSs have been formulated by employing the classical l_p metric, paralleling the advancements in similarity measures. However, in the field of Intuitionistic fuzzy soft sets, no similarity measure has been proposed utilizing the properties of l_p matrices. The reflective symmetry of the l_p metric ensures that similarity measures between sequences of IFSS remain invariant under rearrangements. Furthermore, introducing weights to the elements of l_p-based measures can alter the geometry while preserving its topology. The consideration of such scenarios has revealed unreliable outcomes in computing similarity measures, emphasizing the need for further development.

In response to this, we introduce the concept of intuitionistic fuzzy soft bounded variation (IFSBV) to address distances between IFSSs. IFSBV is employed to approximate the features of IFSSs through the intuitionistic fuzzy soft-valued function (IFSVF). This paper presents a modified distance measure, combining an IFSBV-based measure and a distance measure based on the l_p metric. Consequently, a new similarity measure is developed in this study. The effectiveness of the proposed similarity measure is verified through comprehensive numerical simulations, addressing a clustering problem in software data. Additionally, a thorough comparison with recently reported measures is conducted. The findings suggest that the proposed similarity measure effectively addresses the shortcomings of current approaches.

This paper contributes significantly in the following aspects:

- Proposing a novel intuitionistic fuzzy soft metric by combining IFSBV and the intuitionistic fuzzy soft l_p metric to demonstrate a similarity measure for comparing IFSSs.
- Employing the proposed similarity measure to solve the clustering problem.

- Analyzing and discussing the performance of the modified similarity measure.
- Comparative analysis demonstrating the superior performance of the proposed similarity measure compared to current similarity measures.

The subsequent sections of the paper are organized as follows: In Section 2, we provide mathematical definitions pertinent to our research. Our novel intuitionistic fuzzy soft metric and similarity measure, along with a mathematical demonstration of their properties are demonstrated in Section 3. It also explores the logical reasoning behind the IFSBV-based distance measure between IFSSs. Section 4 addresses a clustering problem, and a comprehensive analysis of the results is carried out, comparing them with established similarity measures. To conclude, Section 5 summarizes the paper and outlines potential avenues for future research.

2. Preliminaries

In this section, we introduce the essential definitions relevant to the current study. We designate W as the universe of discourse, A as the parameter set, and $\widetilde{P}(W)$ represents the power set of W. Additionally, within this context, we have I being a subset of A.

Definition 2.1 ([21])**.** Consider a mapping $S : A \to \widetilde{P}(W)$. A soft set, denoted as (S, A), is defined over the set W

$$(S, A) = \{(S(a), a) : S(a) \in \widetilde{P}(W), a \in A\}$$

Definition 2.2 ([1])**.** A set I defined on the universal set W is termed an intuitionistic fuzzy set when

$$I = \{(w, \mu_I(w), \nu_I(w)) : w \in W\}$$

Here, $\mu_I : W \to [0, 1]$ and $\nu_I : W \to [0, 1]$ represent the membership and non-membership functions, respectively. Additionally, it holds that $0 \le \mu_I(w)+\nu_I(w) \le 1$ for every element $w \in W$.

Definition 2.3 ([22])**.** An intuitionistic fuzzy soft set is represented as the ordered pair (S, I), defined by the relation:

$$(S, I) = \{(\lambda_I(a), a) : \lambda_I(a) \in S(W), a \in A\}.$$

Here, $S(W)$ encompasses all IFSs over W, and λ_I acts as an approximation function given by:

$$\lambda_I : A \to S(W) \text{ such that } \lambda_I(a) = \emptyset \text{ if } a \notin I$$

In this context, the intuitionistic fuzzy empty set is symbolized as $\emptyset$, and $\lambda_I(a)$ is an IFS represented as:

$$\lambda_I(a) = \{(w, \mu_{\lambda_I(a)}(w), \nu_{\lambda_I(a)}(w)) : w \in W\}$$

for all $a \in A$. Moreover, $\mu_{\lambda_I} : W \to [0, 1]$ and $\nu_{\lambda_I} : W \to [0, 1]$ denote the membership and non-membership degree values, respectively, satisfying the condition:

$$0 \le \mu_{\lambda_I(a)}(w) + \nu_{\lambda_I(a)}(w) \le 1$$

for all $w \in W$ to the intuitionistic fuzzy set $\lambda_I(a)$.

Definition 2.4 ([32])**.** An association matrix $\mathcal{T} = (x_{ij})_{m\times m}$ is a matrix in which each entry x_{ij} denotes the association coefficient between S_i and S_j, where $S_i, S_j \in IFSS$.

Definition 2.5 ([32])**.** The operation of composition, denoted as (o), is applied to an association matrix $\mathcal{T}$ and is defined as $\mathcal{T} o\, \mathcal{T} = \mathcal{T}^2 = (x_{ij})$, where each element x_{ij} is determined by $x_{ij} = \max_k \{\min\{x_{ik}, x_{kj}\}\}$ for $i, j = 1, 2, \ldots, , m$. If $\mathcal{T}^2 \subset \mathcal{T}$, indicating that $x_{ij} \geq \max_k \{\min\{x_{ik}, x_{kj}\}\}$ holds for $i, j = 1, 2, \ldots, , m$, then $\mathcal{T}$ is termed an equivalent association matrix.

Definition 2.6 ([32])**.** The matrix $\mathcal{T}_\vartheta = (\vartheta x_{ij})_{m\times m}$ is known as the ϑ-cutting matrix derived from $\mathcal{T}$, where $\mathcal{T}$ is an equivalent association matrix. The computation of each element in $\mathcal{T}_\vartheta$, represented as the $(i,j)^{th}$ element, is determined by the following expression:

$$\vartheta A_{ij} = \begin{cases} 0 & A_{ij} < \vartheta \\ 1 & A_{ij} \geq \vartheta \end{cases} \quad i, j = 1, 2, \ldots, , m$$

Definition 2.7 ([25])**.** Let IFSSs S_1, S_2, and S_3 be defined within W. A metric/distance measure, labelled as $\mathcal{D} : IFSS \times IFSS \to [0,1]$, is a function between IFSSs that satisfies the following criteria:

1. $0 \leq \mathcal{D}(S_1, S_2) \leq 1$.
2. $\mathcal{D}(S_1, S_2) = 0 \Leftrightarrow S_1 = S_2$.
3. $\mathcal{D}(S_1, S_2) = \mathcal{D}(S_2, S_1)$.
4. If $S_1 \subseteq S_2 \subseteq S_3$ then $\mathcal{D}(S_1, S_3) \geq \mathcal{D}(S_1, S_2)$ and $\mathcal{D}(S_1, S_3) \geq \mathcal{D}(S_2, S_3)$.

The relationship between the distance measure and similarity measure is denoted as, $\mathcal{S}(S_1, S_2) = 1 - \mathcal{D}(S_1, S_2)$.

Definition 2.8 ([32])**.** Let's consider two IFSSs, S_1 and S_2, defined in the universe of discourse W. A similarity measure, represented as $\mathcal{S} : IFSS \times IFSS \to [0,1]$, is a function that adheres to the following four conditions:

1. $0 \leq \mathcal{S}(S_1, S_2) \leq 1$.
2. $\mathcal{S}(S_1, S_2) = 1 \Leftrightarrow S_1 = S_2$.
3. $\mathcal{S}(S_1, S_2) = \mathcal{S}(S_2, S_1)$.
4. If $S_1 \subseteq S_2 \subseteq S_3$ then $\mathcal{S}(S_1, S_3) \geq \mathcal{S}(S_1, S_2)$ and $\mathcal{S}(S_1, S_3) \geq \mathcal{S}(S_2, S_3)$.

3. Proposed similarity measure

In classical mathematics, bounded variation (BV) is an essential notion that helps approximate the length of the arc of a real-valued function. Researchers have investigated the theoretical aspects of BV in [24], working within the framework of the FS theory. For fuzzy-valued functions (FVF), Gong and Wu [10] explained the relationship between absolute continuity and BV, and Talo [27] explained the $p-BV$ sequence space of fuzzy numbers (FN). The characteristics of the BV double sequence space of fuzzy real numbers, including solidity, symmetry, and convergence, were examined by Tripathy et al. [28]. Narukawa et al. [23] defined the Choquet integral by using a non-monotonic fuzzy measure of BV to determine the weighted distance of two IFSs. The concept of BV functionality was first introduced by K.

Hirota [11], and effectively applied it to the analysis of pattern recognition. Some essential definitions related to these works are given as follows:

Definition 3.1 ([34])**.** In an interval $[x, y] \subset \mathbb{R}$, a partition n signifies a collection of points $w_0, w_1, \ldots, ., w_n$ where $x = w_0 < w_1 < w_2 ... < w_n = y$.

Definition 3.2 ([7])**.** The definition of p-summable total variation $\mathcal{V}_p$ for a function $\dot{f}$ with respect to the partition n is given below:

$$\mathcal{V}_p(\dot{f}) = \left(|\dot{f}(w_1)|^p + \sum_{i=2}^{n} |\Delta \dot{f}(w_i)|^p \right)^{\frac{1}{p}}$$

where, $\Delta \dot{f}(w_i) = \dot{f}(w_i) - \dot{f}(w_{i-1})$.

Definition 3.3 ([7])**.** If the total variation of a function $\dot{f}$ on a given interval $[x, y]$ is finite, then $\dot{f}$ is said to be a bounded variation (BV) function, meaning that

$$\dot{f} \in BV_p([x, y]) \Leftrightarrow \sup_n \mathcal{V}_p(\dot{f}) < +\infty$$

Definition 3.4 ([31])**.** Score function $\widehat{\alpha} : \Sigma \to [-1, 1]$ is defined as

$$\widehat{\alpha}(w_i) = \mu(w_i) - \nu(w_i).$$

In this context, Σ represents the set comprising all intuitionistic fuzzy values, which can be defined as $\Sigma = \{\langle w, \mu, \nu \rangle | w \in W\}$.

Definition 3.5. An IFSS $S = \{((\mu_S(w_i), \nu_S(w_i)), a_j),\ w_i \in W,\ a \in A,\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m\}$ is said to be a partition of Σ if

$$\widehat{\alpha}(w_1, \mu(w_1), \nu(w_1)) < \widehat{\alpha}(w_2, \mu(w_2), \nu(w_2)) < \cdots < \widehat{\alpha}(w_n \mu(w_n), \nu(w_n)) \text{ [for each parameter } a_j]$$

where $\widehat{\alpha}$ is the score function.

Now, let the partition of Σ, represented by $S_{\widehat{\alpha}} = \{\{(w_i, \mu_i, \nu_i), a_j\} | i = 1, 2, \ldots, n, j = 1, 2, \ldots, m\}$, defines an IFSS in which $\widehat{\alpha}$ is used as a score function as defined in Definition (3.4). For the partition $S_{\widehat{\alpha}}$ and by taking $p \leq 1$, we define the IFSBV bv_p of $\dot{F}_{S_{\widehat{\alpha}}}$ in this way:

$$bv_p(\dot{F}_{S_{\widehat{\alpha}}}) = \frac{1}{2nm} \sum_{j=1}^{m} \Big[\sum_{i=1}^{l} |\dot{F}_{S_{\widehat{\alpha}}}\{(w_i, \mu_i, \nu_i), a_j\}|^p + \sum_{k=l+1}^{n} |\Delta \dot{F}_{S_{\widehat{\alpha}}}\{(w_k, \mu_k, \nu_k), a_j\}|^p \Big]$$
$$< +\infty \tag{3.1}$$

where, $\Delta \dot{F}_{S_{\alpha}}\{(w_k, \mu_k, \nu_k), a_j\} = \dot{F}_{S_{\widehat{\alpha}}}\{(w_k, \mu_k, \nu_k), a_j\} - \dot{F}_{S_{\widehat{\alpha}}}\{(w_{k-1}, \mu_{k-1}, \nu_{k-1}), a_j\}$.

$bv_p(\dot{F}_{S_{\widehat{\alpha}}})$ is used in Equation (3.1) to calculate the lengths of arc of the function $\dot{F}_{S_{\widehat{\alpha}}}$ across the partition $S_{\widehat{\alpha}}$ of the domain Σ. Different values of p in bv_p and different scenarios of $S_{\widehat{\alpha}}$ can be used to characterise the unknown intuitionistic fuzzy soft-valued function (IFSVF) $\dot{F}$. Let us assume two IFSSs as $S = \{((\mu_S(w_i), \nu_S(w_i)), a_j);\ w_i \in W,\ a \in A,\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m\}$ and $T = \{((\mu_T(w_i), \nu_T(w_i)), a_j);\ w_i \in W,\ a \in A,\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m\}$. Hence, $\dot{F}_S$ and $\dot{F}_T$ are considered to be IFSVFs corresponding to the partitions S

and T of the domain Σ. Then, we define the distance measure between S and T as follows:

$$\mathcal{D}_{bv_p}(S,T) = \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\dot{F}_S\{(w_i,\mu_i,\nu_i),a_j\} - \dot{F}_T\{(w_i,\mu_i,\nu_i),a_j\}|^p\Big)$$

$$+\sum_{k=l+1}^{n}\Big(|\Delta\dot{F}_S\{(w_k,\mu_k,\nu_k),a_j\} - \Delta\dot{F}_T\{(w_k,\mu_k,\nu_k),a_j\}|^p\Big)\Big] \tag{3.2}$$

In Equation (3.2), the operator Δ computes the difference between the corresponding elements of the IFSSs. The continuous function $\dot{F}$ over $\dot{F}_S$ and $\dot{F}_T$ is approximated component-wise by the distance measure $\mathcal{D}_{bv_p}(S,T)$ based on IFSBV. The length of $\dot{F}$ is estimated by the IFSBVs $bv_p(\dot{F}_S)$ and $bv_p(\dot{F}_T)$ of the IFSVFs $\dot{F}_S$ and $\dot{F}_T$. The IFSVFs $\dot{F}_S$ and $\dot{F}_T$ are sub-functions of $\dot{F}$ with domain Σ since S and T are partitions of the domain Σ. In summary, for every p value in bv_p, the components of the partitions S and T are positioned on different graphs of the IFSVF $\dot{F}$. To illustrate, let us consider a graph of $\dot{F}$. For each $\{(w_1,\mu_1,\nu_1),a_j\}$, we have $\dot{F}\{(w_1,\mu_1,\nu_1),a_j\}$, meaning that $\dot{F}\{(w_1,\mu_1,\nu_1),a_j\} = \{(w_1,\mu_1,\nu_1),a_j\}$. As a result, the distance generated by IFSBV between S and T is defined as :

$$\mathcal{D}_{bv_p}(S,T) = \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\mu_{S(a_j)}(w_i)-\mu_{T(a_j)}(w_i)|^p+|\nu_{S(a_j)}(w_i)-\nu_{T(a_j)}(w_i)|^p\Big)$$

$$+\sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k)-\Delta\mu_{T(a_j)}(w_k)|^p+|\Delta\nu_{S(a_j)}(w_k)-\Delta\nu_{T(a_j)}(w_k)|^p\Big)\Big] \tag{3.3}$$

Theorem 3.6. *$\mathcal{D}_{bv_p}$ is a distance measure.*

Proof. In order to designate $\mathcal{D}_{bv_p}$ as a metric, it is essential to confirm that it obeys the properties specified in the provided Definition (2.7).

Condition 1. To prove $0 \le \mathcal{D}_{bv_p}(S,T) \le 1$.
Proof. Here,

$$0 \le \sum_{j=1}^{m}\sum_{i=1}^{l}\Big(|\mu_{S(a_j)}(w_i) - \mu_{T(a_j)}(w_i)|^p + |\nu_{S(a_j)}(w_i) - \nu_{T(a_j)}(w_i)|^p\Big) \le 2m \tag{3.4}$$

also,

$$0 \le \sum_{j=1}^{m}\sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k)-\Delta\mu_{T(a_j)}(w_k)|^p+|\Delta\nu_{S(a_j)}(w_k)-\Delta\nu_{T(a_j)}(w_k)|^p\Big)$$

$$\le 2m(n-1) \tag{3.5}$$

Then from eq. (3.4) and (3.5) we have

$$0 \le \sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\mu_{S(a_j)}(w_i) - \mu_{T(a_j)}(w_i)|^p + |\nu_{S(a_j)}(w_i) - \nu_{T(a_j)}(w_i)|^p\Big)$$

$$+\sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k)\Delta\mu_{T(a_j)}(w_k)|^p+|\Delta\nu_{S(a_j)}(w_k)-\Delta\nu_{T(a_j)}(w_k)|^p\Big)\Big]$$
$$\leq 2m+2m(n-1)=2nm$$
$$\Rightarrow 0\leq\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\mu_{S(a_j)}(w_i)-\mu_{T(a_j)}(w_i)|^p+|\nu_{S(a_j)}(w_i)-\nu_{T(a_j)}(w_i)|^p\Big)$$
$$+\sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k)\Delta\mu_{T(a_j)}(w_k)|^p+|\Delta\nu_{S(a_j)}(w_k)-\Delta\nu_{T(a_j)}(w_k)|^p\Big)\Big]$$
$$\leq 1$$

Therefore, $0\leq\mathcal{D}_{bv_p}(S,T)\leq 1$.

Condition 2. To prove $\mathcal{D}_{bv_p}(S,T)=0\Leftrightarrow S=T$.
Proof. Let $\mathcal{D}_{bv_p}(S,T)=0$

$$\Leftrightarrow \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\mu_{S(a_j)}(w_i)-\mu_{T(a_j)}(w_i)|^p+|\nu_{S(a_j)}(w_i)-\nu_{T(a_j)}(w_i)|^p\Big)$$
$$+\sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k)\Delta\mu_{T(a_j)}(w_k)|^p+|\Delta\nu_{S(a_j)}(w_k)-\Delta\nu_{T(a_j)}(w_k)|^p\Big)\Big]$$
$$=0$$

Therefore,

$$|\mu_{S(a_j)}(w_i)-\mu_{T(a_j)}(w_i)|=0\Leftrightarrow\mu_{S(a_j)}(w_i)=\mu_{T(a_j)}(w_i)\tag{3.6}$$

$$|\nu_{S(a_j)}(w_i)-\nu_{T(a_j)}(w_i)|=0\Leftrightarrow\nu_{S(a_j)}(w_i)=\nu_{T(a_j)}(w_i)\tag{3.7}$$

$$|\Delta\mu_{S(a_j)}(w_k)-\Delta\mu_{T(a_j)}(w_k)|=0\tag{3.8}$$

$$|\Delta\nu_{S(a_j)}(w_k)-\Delta\nu_{T(a_j)}(w_k)|=0\tag{3.9}$$

From equation (3.8) we have,

$$\mu_{S(a_j)}(w_{k+1})-\mu_{S(a_j)}(w_k)-\mu_{T(a_j)}(w_{k+1})+\mu_{T(a_j)}(w_k)=0$$

Now, suppose $k=1$ then from eq. (3.6) we have, $\mu_{S(a_j)}(w_1)=\mu_{T(a_j)}(w_1)$. Then by using mathematical induction we can easily proof that $\mu_{S(a_j)}(w_k)=\mu_{T(a_j)}(w_k)$ and hence,

$$\mu_{S(a_j)}(w_{k+1})-\mu_{T(a_j)}(w_{k+1})=0\Leftrightarrow\mu_{S(a_j)}(w_{k+1})=\mu_{T(a_j)}(w_{k+1})\tag{3.10}$$

Similarly, From Equation (7) we have,

$$\nu_{S(a_j)}(w_{k+1}) - \nu_{T(a_j)}(w_{k+1}) = 0 \Leftrightarrow \nu_{S(a_j)}(w_{k+1}) = \nu_{T(a_j)}(w_{k+1}) \tag{3.11}$$

Now, we can conclude that for all $i \in \mathbb{N}$,

$$\mathcal{D}_{bv_p}(S,T) = 0 \quad \Leftrightarrow \quad \mu_{S(a_j)}(w_i) = \mu_{T(a_j)}(w_i), \quad \nu_{S(a_j)}(w_i) = \nu_{T(a_j)}(w_i) \quad \Leftrightarrow S = T$$

Condition 3. To prove $\mathcal{D}_{bv_p}(S,T) = \mathcal{D}_{bv_p}(T,S)$.
Proof. It is straightforward to demonstrate for any IFSSs that, from the definition of $\mathcal{D}_{bv_p}$

$$\mathcal{D}_{bv_p}(S,T) = \mathcal{D}_{bv_p}(T,S)$$

Condition 4. To prove, if $S \subseteq T \subseteq R$, then $\mathcal{D}_{bv_p}(S,R) \geq \mathcal{D}_{bv_p}(S,T)$. and $\mathcal{D}_{bv_p}(S,R) \geq \mathcal{D}_{bv_p}(T,R)$.
Proof. In order to prove the fourth requirement, it is same as demonstrating transitivity. Assuming that three IFSSs are given as S, T, and, R, then

$$\begin{aligned}
\mathcal{D}_{bv_p}(S,R) &= \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}(|\mu_{S(a_j)}(w_i) - \mu_{R(a_j)}(w_i)|^p + |\nu_{S(a_j)}(w_i) - \nu_{R(a_j)}(w_i)|^p) \\
&\quad + \sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k) - \Delta\mu_{R(a_j)}(w_k)|^p + |\Delta\nu_{S(a_j)}(w_k) - \Delta\nu_{R(a_j)}(w_k)|^p\Big)\Big] \\
&= \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}(|\mu_{S(a_j)}(w_i) - \mu_{T(a_j)}(w_i) + \mu_{T(a_j)}(w_i) - \mu_{R(a_j)}(w_i)|^p \\
&\quad + |\nu_{S(a_j)}(w_i) - \nu_{T(a_j)}(w_i) + \nu_{T(a_j)}(w_i) - \nu_{R(a_j)}(w_i)|^p) \\
&\quad + \sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k) - \Delta\mu_{T(a_j)}(w_k) + \Delta\mu_{T(a_j)}(w_k) - \Delta\mu_{R(a_j)}(w_k)|^p \\
&\quad + |\Delta\nu_{S(a_j)}(w_k) - \Delta\nu_{T(a_j)}(w_k) + \Delta\nu_{T(a_j)}(w_k) - \Delta\nu_{R(a_j)}(w_k)|^p\Big)\Big] \\
&\leq \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\mu_{S(a_j)}(w_i) - \mu_{T(a_j)}(w_i)|^p + |\mu_{T(a_j)}(w_i) - \mu_{R(a_j)}(w_i)|^p \\
&\quad + |\nu_{S(a_j)}(w_i)\nu_{T(a_j)}(w_i)|^p + |\nu_{T(a_j)}(w_i) - \nu_{R(a_j)}(w_i)|^p\Big) \\
&\quad + \sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k) - \Delta\mu_{T(a_j)}(w_k)|^p + |\Delta\mu_{T(a_j)}(w_k) - \Delta\mu_{R(a_j)}(w_k)|^p \\
&\quad + |\Delta\nu_{S(a_j)}(w_k) - \Delta\nu_{T(a_j)}(w_k)|^p + |\Delta\nu_{T(a_j)}(w_k) - \Delta\nu_{R(a_j)}(w_k)|^p\Big)\Big]
\end{aligned}$$

[By triangle inequality]

$\mathcal{D}_{bv_p}(S,R) \leq \mathcal{D}_{bv_p}(S,T) + \mathcal{D}_{bv_p}(T,R)$

Therefore, $\mathcal{D}_{bv_p}$ satisfies all the conditions of metric. □

Theorem 3.7. *The proposed metric is normable.*

Proof. To proof this theorem we need to show that $\mathcal{D}_{bv_p}$ satisfies translation invariant and homogeneity conditions.

1. Translation invariant: Let,

$$S = \{(\mu_S(w_i), \nu_S(w_i)), a_j\}$$
$$T = \{(\mu_T(w_i), \nu_T(w_i)), a_j\}$$
$$R = \{(\mu_R(w_i), \nu_R(w_i)), a_j\}$$

Then,

$$S + T = \{(\mu_S(w_i) + \mu_T(w_i), \nu_S(w_i) + \mu_T(w_i)), a_j\}$$
$$T + R = \{(\mu_T(w_i) + \mu_R(w_i), \nu_T(w_i) + \mu_R(w_i)), a_j\}$$

Therefore,

$$\begin{aligned}
&\mathcal{D}_{bv_p}(S+T, T+R) \\
&= \frac{1}{2nm} \sum_{j=1}^{m} \Big[\sum_{i=1}^{l} \Big(|\mu_{S(a_j)}(w_i) + \mu_{T(a_j)}(w_i) - \mu_{T(a_j)}(w_i) - \mu_{R(a_j)}(w_i)|^p \\
&\quad + |\nu_{S(a_j)}(w_i) + \nu_{T(a_j)}(w_i) - \nu_{T(a_j)}(w_i) - \nu_{R(a_j)}(w_i)|^p \Big) \\
&\quad + \sum_{k=l+1}^{n} \Big(|\Delta\mu_{S(a_j)}(w_k) + \Delta\mu_{T(a_j)}(w_k) - \Delta\mu_{T(a_j)}(w_k) - \Delta\mu_{R(a_j)}(w_k)|^p \\
&\quad + |\Delta\nu_{S(a_j)}(w_k) + \Delta\nu_{T(a_j)}(w_k) - \Delta\nu_{T(a_j)}(w_k) - \Delta\nu_{R(a_j)}(w_k)|^p \Big) \Big] \\
&= \frac{1}{2nm} \sum_{j=1}^{m} \Big[\sum_{i=1}^{l} \Big(|\mu_{S(a_j)}(w_i) - \mu_{R(a_j)}(w_i)|^p + |\nu_{S(a_j)}(w_i) - \nu_{R(a_j)}(w_i)|^p \Big) \\
&\quad + \sum_{k=l+1}^{n} \Big(|\Delta\mu_{S(a_j)}(w_k) - \Delta\mu_{R(a_j)}(w_k)|^p + |\Delta\nu_{S(a_j)}(w_k) - \Delta\nu_{R(a_j)}(w_k)|^p \Big) \Big] \\
&\therefore \mathcal{D}_{bv_p}(S+T, T+R) = \mathcal{D}_{bv_p}(S,R)
\end{aligned}$$

2. Homogeneity: To prove, $\mathcal{D}_{bv_p}(\kappa S, \kappa T) = \kappa \mathcal{D}_{bv_p}(S,T)$.
Here,

$$\begin{aligned}
&\mathcal{D}_{bv_p}(\kappa S, \kappa T) \\
&= \frac{1}{2nm} \sum_{j=1}^{m} \Big[\sum_{i=1}^{l} \Big(|\kappa\mu_{S(a_j)}(w_i) - \kappa\mu_{T(a_j)}(w_i)|^p + |\kappa\nu_{S(a_j)}(w_i) - \kappa\nu_{T(a_j)}(w_i)|^p \Big) \\
&\quad + \sum_{k=l+1}^{n} \Big(|\Delta(\kappa\mu_{S(a_j)}(w_k)) - \Delta(\kappa\mu_{T(a_j)}(w_k))|^p
\end{aligned}$$

$$+|\Delta(\kappa\nu_{S(a_j)}(w_k)) - \Delta(\kappa\nu_{T(a_j)}(w_k))|^p\Big)\Big] \tag{3.12}$$

Here,

$$\begin{aligned}\Delta(\kappa\mu_{S(a_j)}(w_k)) &= (\kappa\mu_{S(a_j)}(w_{k+1}) - \kappa\mu_{S(a_j)}(w_k))\\ &= \kappa(\mu_{S(a_j)}(w_{k+1}) - \mu_{S(a_j)}(w_k))\\ &= \kappa\Delta(\mu_{S(a_j)}(w_k)) \end{aligned}\tag{3.13}$$

Similarly,

$$\Delta(\kappa\mu_{T(a_j)}(w_k)) = \kappa\Delta(\mu_{T(a_j)}(w_k)) \tag{3.14}$$

$$\Delta(\kappa\nu_{S(a_j)}(w_k)) = \kappa\Delta(\nu_{S(a_j)}(w_k)) \tag{3.15}$$

$$\Delta(\kappa\nu_{T(a_j)}(w_k)) = \kappa\Delta(\nu_{T(a_j)}(w_k)) \tag{3.16}$$

Then from Eq. (3.18) we have,

$$\begin{aligned}&\mathcal{D}_{bv_p}(\kappa S, \kappa T)\\ &= \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\kappa||\mu_{S(a_j)}(w_i) - \mu_{T(a_j)}(w_i)|^p + |\kappa||\nu_{S(a_j)}(w_i) - \nu_{T(a_j)}(w_i)|^p\Big)\\ &\quad + \sum_{k=l+1}^{n}\Big(|\kappa||\Delta(\mu_{S(a_j)}(w_k)) - \Delta(\mu_{T(a_j)}(w_k))|^p\\ &\quad +|\kappa||\Delta(\nu_{S(a_j)}(w_k)) - \Delta(\nu_{T(a_j)}(w_k))|^p\Big)\Big]\\ &= |\kappa|\frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{i=1}^{l}\Big(|\mu_{S(a_j)}(w_i) - \mu_{T(a_j)}(w_i)|^p + |\nu_{S(a_j)}(w_i) - \nu_{T(a_j)}(w_i)|^p\Big)\\ &\quad + \sum_{k=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_k) - \Delta\mu_{T(a_j)}(w_k)|^p + |\Delta\nu_{S(a_j)}(w_k) - \Delta\nu_{T(a_j)}(w_k)|^p\Big)\Big]\\ &= |\kappa|\mathcal{D}_{bv_p}(S,T)\end{aligned}$$

Hence proved.

□

Eq. (3.3) provides the metric, which we can use to evaluate the similarity measures between IFSSs. This is because metric/distance and similarity measures are dual concepts in terms of fuzzy complement. Therefore,

$$\begin{aligned}\mathcal{S}_{bv_p}(S,T) &= 1 - \mathcal{D}_{bv_p}(S,T)\\ &= 1 - \frac{1}{2nm}\sum_{j=1}^{m}\Big[\sum_{k=1}^{l}\Big(|\mu_{S(a_j)}(w_k) - \mu_{T(a_j)}(w_k)|^p\\ &\quad + |\nu_{S(a_j)}(w_k) - \nu_{T(a_j)}(w_k)|^p\Big)\\ &\quad + \sum_{i=l+1}^{n}\Big(|\Delta\mu_{S(a_j)}(w_i) - \Delta\mu_{T(a_j)}(w_i)|^p + |\Delta\nu_{S(a_j)}(w_i) - \Delta\nu_{T(a_j)}(w_i)|^p\Big)\Big]\end{aligned}\tag{3.17}$$

4. IFSS-based clustering

Clustering serves as a technique for partitioning data into discrete clusters or groups and similarity measures hold pivotal importance in tackling clustering problems ([17,32]). The efficacy of clustering methods is heavily reliant on these similarity measures, as the elements that show higher similarity values ideally find placement within the same cluster. Xu introduced the Intuitionistic Fuzzy Set based Clustering (IFSC) algorithm, which is an influential algorithm in fuzzy clustering, in his work [32]. This algorithm has been extensively used in a variety of clustering problems. In a similar vein, a pythogonal fuzzy soft set clustering algorithm has been proposed in [2]. This paper aims to present an approach termed Intuitionistic Fuzzy Soft Set based Clustering (IFSSC), wherein this clustering algorithm is seamlessly integrated into a hierarchical fuzzy clustering technique by incorporating the proposed similarity measure. The IFSSC algorithm arranges data points into a hierarchical structure resembling a tree, organized according to the extent of their similarity.

We have substituted the initial similarity measure in the IFSSC algorithm with our newly developed modified similarity measure ($\mathcal{S}_{bv_p}$), and we show its potential for use in clustering problems. Algorithm-1 provides a detailed, step-by-step technique that outlines the whole IFSSC algorithm. A detailed description of the clustering problems presented in [2] is provided in the part that follows. We have utilised the innovative IFSSC algorithm to tackle these problems.

4.0.1. *Clustering Example.* We use the dataset given in [2], which consists of 10 datasets (Table 2), to run the clustering algorithm. a_1, a_2, a_3, a_4, and a_5 are the five parameters that make up the data set. Each dataset is categorised based on the following criteria: $w_1 = Image\ processing$, $w_2 = Measurement\ equipment$, $w_3 = Digital\ surface\ models$, and $w_4 = Production\ of\ 3D\ modelling$. By considering $p = 1$, the steps of the IFSSC algorithm are as follows:

Algorithm 1 Proposed IFSSC algorithm

Input: Provided a number of IFSSs represented by the notation $\{S_1, S_2, \ldots, S_n\}$.
Output: Results of clustering are correlated with different levels of confidence.

1: Compute the association coefficients $x_{i,j}$ using the similarity measure $\mathcal{S}_{bv_p}$ for all pairings (S_i, S_j), where $i, j = 1, 2, \ldots, n$.
2: Calculate the association matrix $\mathcal{T}$ using the values of $x_{i,j}$.
3: **If** $\mathcal{T}$ is considered to be an equivalent association matrix according to the Definition 2.4.
4: Determine the ϑ-cutting matrix by using Definition 2.6 to determine the confidence level ϑ.
5: **Else**
6: Determine the equivalence of $\mathcal{T}$ by using the composition operation ($\circ$) as given in Definition 2.5.
7: **End**
8: Get the final resulting equivalent association matrix $\mathcal{T}$.
9: Set up clusters with different levels of confidence. ϑ.

Step 1. The association coefficients are computed and the association coefficients matrix $\mathcal{T}$ is constructed based on the provided data set, taking into account the proposed similarity measure $\mathcal{S}_{bv_p}$.

$$\mathcal{T} = \begin{pmatrix}
1.0000 & 0.9107 & 0.8641 & 0.8444 & 0.8566 & 0.8012 & 0.8578 & 0.8076 & 0.7671 & 0.7692 \\
0.9107 & 1.0000 & 0.8538 & 0.8277 & 0.8326 & 0.7970 & 0.8341 & 0.8055 & 0.7661 & 0.7468 \\
0.8641 & 0.8538 & 1.0000 & 0.8092 & 0.8056 & 0.7670 & 0.8156 & 0.7803 & 0.7833 & 0.7596 \\
0.8444 & 0.8277 & 0.8092 & 1.0000 & 0.8388 & 0.7719 & 0.8609 & 0.8000 & 0.7604 & 0.8107 \\
0.8566 & 0.8326 & 0.8056 & 0.8388 & 1.0000 & 0.7748 & 0.8278 & 0.7522 & 0.7552 & 0.7712 \\
0.8012 & 0.7970 & 0.7670 & 0.7719 & 0.7748 & 1.0000 & 0.7647 & 0.8231 & 0.7456 & 0.7523 \\
0.8578 & 0.8341 & 0.8156 & 0.8609 & 0.8278 & 0.7647 & 1.0000 & 0.7542 & 0.7260 & 0.7862 \\
0.8076 & 0.8055 & 0.7803 & 0.8000 & 0.7522 & 0.8231 & 0.7542 & 1.0000 & 0.7941 & 0.7597 \\
0.7671 & 0.7661 & 0.7833 & 0.7604 & 0.7552 & 0.7456 & 0.7260 & 0.7941 & 1.0000 & 0.7884 \\
0.7692 & 0.7468 & 0.7596 & 0.8107 & 0.7712 & 0.7523 & 0.7862 & 0.7597 & 0.7884 & 1.0000
\end{pmatrix}$$

Step 2. We must compute $\mathcal{T}^{2^j}$ for $j = 2, 3, \ldots$ in order to ascertain the equivalence of $\mathcal{T}$. According to Definition (2.5), $\mathcal{T}^2 = \mathcal{T} o \mathcal{T}$. We can evaluate whether or not $\mathcal{T}$ satisfies the equivalence condition with this computation.

$$\mathcal{T}^2 = \begin{pmatrix}
1.0000 & 0.9107 & 0.8641 & 0.8578 & 0.8566 & 0.8076 & 0.8578 & 0.8076 & 0.7941 & 0.8107 \\
0.9107 & 1.0000 & 0.8641 & 0.8578 & 0.8566 & 0.8076 & 0.8578 & 0.8076 & 0.7941 & 0.8107 \\
0.8641 & 0.8641 & 1.0000 & 0.8578 & 0.8566 & 0.8076 & 0.8578 & 0.8076 & 0.7941 & 0.8107 \\
0.8578 & 0.8578 & 0.8578 & 1.0000 & 0.8566 & 0.8076 & 0.8609 & 0.8076 & 0.7941 & 0.8107 \\
0.8566 & 0.8566 & 0.8566 & 0.8566 & 1.0000 & 0.8076 & 0.8566 & 0.8076 & 0.7941 & 0.8107 \\
0.8076 & 0.8076 & 0.8076 & 0.8076 & 0.8076 & 1.0000 & 0.8076 & 0.8231 & 0.7941 & 0.8076 \\
0.8578 & 0.8578 & 0.8578 & 0.8609 & 0.8566 & 0.8076 & 1.0000 & 0.8076 & 0.7941 & 0.8107 \\
0.8076 & 0.8076 & 0.8076 & 0.8076 & 0.8076 & 0.8231 & 0.8076 & 1.0000 & 0.7941 & 0.8076 \\
0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 1.0000 & 0.7941 \\
0.8107 & 0.8107 & 0.8107 & 0.8107 & 0.8107 & 0.8076 & 0.8107 & 0.8076 & 0.7941 & 1.0000
\end{pmatrix}$$

Step 3. In this case, $\mathcal{T}^2$ is not an equivalent matrix; thus, we must keep computing $\mathcal{T}^{2^j}$ for increasing values of j (e.g., $j = 2, 3, \ldots$) until it's equivalent.

$$\mathcal{T}^4 = \begin{pmatrix}
1.0000 & 0.9107 & 0.8641 & 0.8578 & 0.8566 & 0.8076 & 0.8578 & 0.8076 & 0.7941 & 0.8107 \\
0.9107 & 1.0000 & 0.8641 & 0.8578 & 0.8566 & 0.8076 & 0.8578 & 0.8076 & 0.7941 & 0.8107 \\
0.8641 & 0.8641 & 1.0000 & 0.8578 & 0.8566 & 0.8076 & 0.8578 & 0.8076 & 0.7941 & 0.8107 \\
0.8578 & 0.8578 & 0.8578 & 1.0000 & 0.8566 & 0.8076 & 0.8609 & 0.8076 & 0.7941 & 0.8107 \\
0.8566 & 0.8566 & 0.8566 & 0.8566 & 1.0000 & 0.8076 & 0.8566 & 0.8076 & 0.7941 & 0.8107 \\
0.8076 & 0.8076 & 0.8076 & 0.8076 & 0.8076 & 1.0000 & 0.8076 & 0.8231 & 0.7941 & 0.8076 \\
0.8578 & 0.8578 & 0.8578 & 0.8609 & 0.8566 & 0.8076 & 1.0000 & 0.8076 & 0.7941 & 0.8107 \\
0.8076 & 0.8076 & 0.8076 & 0.8076 & 0.8076 & 0.8231 & 0.8076 & 1.0000 & 0.7941 & 0.8076 \\
0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 0.7941 & 1.0000 & 0.7941 \\
0.8107 & 0.8107 & 0.8107 & 0.8107 & 0.8107 & 0.8076 & 0.8107 & 0.8076 & 0.7941 & 1.0000
\end{pmatrix}$$

Step 4. As can be observed, $\mathcal{T}^2$ is an equivalent association matrix since $\mathcal{T}^2 = \mathcal{T}^4$.

Step 5. To illustrate the relationships between components in the equivalent matrix $\mathcal{T}^2$, the obtained clusters for the classification problem on the ten provided data sets can be derived by selecting different confidence levels ϑ.

Table 1. The obtained clusters of the data sets for various confidence levels ϑ as generated by the Improved IFSSC algorithm

Similarity level with respect to ϑ	Clusters	Number of classes
$0.0000 \leq \vartheta \leq 0.7941$	$\{1,2,3,4,5,6,7,8,10,9\}$	1
$0.7941 < \vartheta \leq 0.8076$	$\{1,2,3,4,5,6,7,8,10\},\{9\}$	2
$0.8076 < \vartheta \leq 0.8107$	$\{1,2,3,4,5,7,10\},\{6,8\},\{9\}$	3
$0.8107 < \vartheta \leq 0.8231$	$\{1,2,3,4,5,7\},\{6\},\{8\},\{9\},\{10\}$	5
$0.8231 < \vartheta \leq 0.8566$	$\{1,2,3,4,5,7\},\{6\},\{8\},\{9\},\{10\}$	5
$0.8566 < \vartheta \leq 0.8578$	$\{1,2,3\},\{4,7\},\{5\},\{6\},\{8\},\{9\},\{10\}$	7
$0.8578 < \vartheta \leq 0.8690$	$\{1,2,3\},\{4,7\},\{5\},\{6\},\{8\},\{9\},\{10\}$	7
$0.8690 < \vartheta \leq 0.8641$	$\{1,2,3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\}$	8
$0.8641 < \vartheta \leq 0.9107$	$\{1,2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\}$	9
$0.9107 < \vartheta \leq 1.0000$	$\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\}$	10

Table 2

Data-1

	a_1	a_2	a_3	a_4	a_5
w_1	(0.481, 0.392)	(0.890, 0.200)	(0.783, 0.152)	(0.872, 0.272)	(0.513, 0.581)
w_2	(0.690, 0.211)	(0.672, 0.312)	(0.662, 0.282)	(0.422, 0.591)	(0.661, 0.041)
w_3	(0.772, 0.233)	(0.691, 0.481)	(0.672, 0.411)	(0.823, 0.513)	(0.691, 0.231)
w_4	(0.781, 0.240)	(0.611, 0.352)	(0.920, 0.290)	(0.740, 0.060)	(0.760, 0.380)

Data-2

	a_1	a_2	a_3	a_4	a_5
w_1	(0.586, 0.318)	(0.890, 0.140)	(0.881, 0.172)	(0.770, 0.450)	(0.572, 0.573)
w_2	(0.662, 0.332)	(0.910, 0.290)	(0.702, 0.261)	(0.271, 0.682)	(0.773, 0.133)
w_3	(0.471, 0.481)	(0.760, 0.550)	(0.661, 0.321)	(0.721, 0.561)	(0.741, 0.234)
w_4	(0.371, 0.542)	(0.640, 0.530)	(0.873, 0.143)	(0.822, 0.292)	(0.774, 0.264)

Data-3

	a_1	a_2	a_3	a_4	a_5
w_1	(0.251, 0.451)	(0.890, 0.011)	(0.754, 0.254)	(0.778, 0.144)	(0.471, 0.672)
w_2	(0.512, 0.321)	(0.641, 0.442)	(0.781, 0.222)	(0.343, 0.853)	(0.442, 0.031)
w_3	(0.322, 0.543)	(0.724, 0.342)	(0.782, 0.561)	(0.872, 0.591)	(0.713, 0.343)
w_4	(0.450, 0.331)	(0.452, 0.442)	(0.889, 0.332)	(0.521, 0.311)	(0.783, 0.371)

Data-4

	a_1	a_2	a_3	a_4	a_5
w_1	(0.342, 0.272)	(0.881, 0.000)	(0.811, 0.156)	(0.571, 0.572)	(0.554, 0.234)
w_2	(0.551, 0.241)	(0.570, 0.360)	(0.783, 0.223)	(0.633, 0.173)	(0.714, 0.312)
w_3	(0.563, 0.233)	(0.730, 0.450)	(0.891, 0.452)	(0.771, 0.441)	(0.892, 0.342)
w_4	(0.713, 0.291)	(0.880, 0.441)	(0.885, 0.465)	(0.662, 0.233)	(0.662, 0.142)

Data-5

	a_1	a_2	a_3	a_4	a_5
w_1	(0.761, 0.331)	(0.783, 0.191)	(0.680, 0.272)	(0.560, 0.150)	(0.514, 0.592)
w_2	(0.322, 0.682)	(0.681, 0.352)	(0.750, 0.262)	(0.570, 0.271)	(0.622, 0.562)
w_3	(0.551, 0.651)	(0.666, 0.546)	(0.752, 0.431)	(0.781, 0.654)	(0.721, 0.322)
w_4	(0.273, 0.673)	(0.462, 0.600)	(0.891, 0.289)	(0.554, 0.252)	(0.773, 0.363)

Data-6

	a_1	a_2	a_3	a_4	a_5
w_1	(0.721, 0.321)	(0.721, 0.244)	(0.520, 0.330)	(0.780, 0.372)	(0.882, 0.121)
w_2	(0.445, 0.274)	(0.381, 0.624)	(0.460, 0.770)	(0.332, 0.672)	(0.881, 0.214)
w_3	(0.672, 0.322)	(0.574, 0.567)	(0.670, 0.540)	(0.822, 0.472)	(0.784, 0.234)
w_4	(0.401, 0.301)	(0.213, 0.733)	(0.440, 0.620)	(0.722, 0.132)	(0.652, 0.652)

Data-7

	a_1	a_2	a_3	a_4	a_5
w_1	(0.621, 0.261)	(0.651, 0.251)	(0.681, 0.162)	(0.670, 0.240)	(0.520, 0.120)
w_2	(0.571, 0.292)	(0.732, 0.362)	(0.892, 0.272)	(0.860, 0.321)	(0.610, 0.240)
w_3	(0.522, 0.362)	(0.772, 0.361)	(0.661, 0.321)	(0.780, 0.330)	(0.640, 0.460)
w_4	(0.722, 0.131)	(0.731, 0.231)	(0.751, 0.142)	(0.660, 0.130)	(0.870, 0.230)

Data-8

	a_1	a_2	a_3	a_4	a_5
w_1	(0.241, 0.481)	(0.881, 0.242)	(0.871, 0.132)	(0.661, 0.471)	(0.552, 0.732)
w_2	(0.641, 0.352)	(0.262, 0.573)	(0.342, 0.783)	(0.232, 0.712)	(0.561, 0.461)
w_3	(0.561, 0.122)	(0.567, 0.611)	(0.573, 0.573)	(0.342, 0.621)	(0.531, 0.622)
w_4	(0.591, 0.311)	(0.212, 0.661)	(0.442, 0.722)	(0.321, 0.711)	(0.212, 0.782)

Data-9

	a_1	a_2	a_3	a_4	a_5
w_1	(0.574, 0.174)	(0.821, 0.121)	(0.820, 0.001)	(0.780, 0.330)	(0.131, 0.671)
w_2	(0.554, 0.060)	(0.212, 0.782)	(0.770, 0.120)	(0.330, 0.780)	(0.330, 0.780)
w_3	(0.550, 0.190)	(0.661, 0.231)	(0.770, 0.140)	(0.540, 0.720)	(0.560, 0.680)
w_4	(0.420, 0.420)	(0.232, 0.872)	(0.890, 0.230)	(0.110, 0.780)	(0.321, 0.771)

Data-10

	a_1	a_2	a_3	a_4	a_5
w_1	(0.450, 0.230)	(0.830, 0.120)	(0.890, 0.230)	(0.560, 0.170)	(0.470, 0.660)
w_2	(0.510, 0.110)	(0.330, 0.820)	(0.730, 0.340)	(0.780, 0.120)	(0.670, 0.210)
w_3	(0.560, 0.260)	(0.760, 0.320)	(0.890, 0.300)	(0.880, 0.420)	(0.660, 0.220)
w_4	(0.880, 0.130)	(0.450, 0.870)	(0.170, 0.820)	(0.770, 0.120)	(0.780, 0.110)

Table 1 presents the clustering results for different confidence levels ϑ, illustrating the clusters that were generated for the data sets. Further, based on Table 1, we perform a sensitivity analysis to observe how various confidence levels ϑ impact the resulting clusters for the given data as follows:

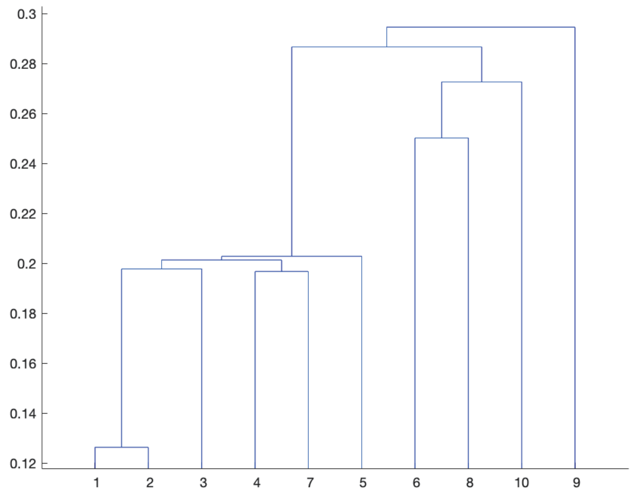

FIGURE 1. Hierarchical clustering tree through the proposed IFSSC algorithm.

(a) If $0.0000 \leq \vartheta \leq 0.7941$, all the given data sets belong to a single cluster, such as, $\mathcal{C}_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

(b) If $0.7941 < \vartheta \leq 0.8076$, then the data sets are categorized into two different types of clusters same as before, such as, $\mathcal{C}_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$ and $\mathcal{C}_2 = \{9\}$.

(c) If $0.8076 < \vartheta \leq 0.8107$, then the data sets are categorized into three different types, such as, $\mathcal{C}_1 = \{1, 2, 3, 4, 5, 7, 10\}$, $\mathcal{C}_2 = \{6, 8\}$, and $\mathcal{C}_3 = \{9\}$ clusters.

(d) If $0.8107 < \vartheta \leq 0.8231$ and $0.8231 < \vartheta \leq 0.8566$, then the data sets are categorized into five different types, such as, $\mathcal{C}_1 = \{1, 2, 3, 4, 5, 7\}$, $\mathcal{C}_2 = \{6\}$, $\mathcal{C}_3 = \{8\}$, $\mathcal{C}_4 = \{9\}$, and $\mathcal{C}_5 = \{10\}$ clusters.

(e) If $0.8566 < \vartheta \leq 0.8578$ and $0.8578 < \vartheta \leq 0.8690$, then the data sets are categorized into seven different types, such as, $\mathcal{C}_1 = \{1, 2, 3\}$, $\mathcal{C}_2 = \{4, 7\}$, $\mathcal{C}_3 = \{5\}$, $\mathcal{C}_4 = \{6\}$, , $\mathcal{C}_5 = \{8\}$, $\mathcal{C}_6 = \{9\}$, and $\mathcal{C}_7 = \{10\}$ clusters.

(f) If $0.8690 < \vartheta \leq 0.8641$, then the data sets are categorized into eight different types, such as, $\mathcal{C}_1 = \{1, 2, 3\}$, $\mathcal{C}_2 = \{4\}$, $\mathcal{C}_3 = \{5\}$, $\mathcal{C}_4 = \{6\}$, , $\mathcal{C}_5 = \{7\}$, $\mathcal{C}_6 = \{8\}$, $\mathcal{C}_7 = \{9\}$, and $\mathcal{C}_8 = \{10\}$ clusters.

(g) If $0.8641 < \vartheta \leq 0.9107$, then the data sets are categorized into nine different types, such as, $\mathcal{C}_1 = \{1, 2\}$, $\mathcal{C}_2 = \{3\}$, $\mathcal{C}_3 = \{4\}$, $\mathcal{C}_4 = \{5\}$, , $\mathcal{C}_5 = \{6\}$, $\mathcal{C}_6 = \{7\}$, $\mathcal{C}_7 = \{8\}$, $\mathcal{C}_8 = \{9\}$, and $\mathcal{C}_9 = \{10\}$ clusters.

(h) If $0.9107 < \vartheta \leq 1.0000$, then the data sets are categorized into ten different types, such as, $\mathcal{C}_1 = \{1\}$, $\mathcal{C}_2 = \{2\}$, $\mathcal{C}_3 = \{3\}$, $\mathcal{C}_4 = \{4\}$, $\mathcal{C}_5 = \{5\}$, $\mathcal{C}_6 = \{6\}$, $\mathcal{C}_7 = \{7\}$, $\mathcal{C}_8 = \{8\}$, $\mathcal{C}_9 = \{9\}$, and $\mathcal{C}_{10} = \{10\}$ clusters.

Figure 1 shows the hierarchical clustering diagram for the equivalent matrix $\mathcal{T}^2$, which was obtained using our proposed IFSSC algorithm using $\mathcal{S}_{bv_p}$ on the IF-SSs data sets. It is clear from looking at the data in Table 1 and Figure 1 that our improved IFSSC algorithm with $\mathcal{S}_{bv_p}$ classifies the IFSS data sets. The resulting clusters are consistent with the clustering results presented in the paper [2]. For example, the results with five clusters $\{1,2,3,4,5,7\},\{6\},\{8\},\{9\},\{10\}$, eight clusters $\{1,2,3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\}$ and the results with nine clusters $\{1,2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\}$ are same as the clusters of [2]. On the other hand, our suggested method is more efficient than the current one since our clustering process terminates at $\mathcal{T}^2$, while the clustering process in [2] terminates at $\mathcal{T}^8$.

The effectiveness of the suggested similarity measure is evident in its performance for the clustering problem mentioned above. However, a notable limitation of our proposed measure is the need to select the input parameter l. Typically, setting $l = \frac{\text{Number of data}}{2}$ proves effective. Therefore, further refinement of this measure, specifically in tuning this parameter, is necessary for optimal performance in handling any type of dataset.

5. Conclusion

In this study, a modified metric/distance measure, referred to as the novel intuitionistic fuzzy soft metric to demonstrate a similarity measure, was introduced for IFSSs. This metric was developed by modifying the existing measure based on IFSBV and the intuitionistic fuzzy soft l_p metric. The resulting modified similarity/distance measures were demonstrated to meet essential mathematical criteria and display comprehensive connections with both IFSBV and l_p metric measures in diverse situations. Moreover, a comparative demonstration has been shown in a clustering problem of software data, with other existing measures. The proposed measure outperforms the existing measures in terms of classification accuracy for clustering problems.

In our future endeavour, we aspire to extend the utilization of the fuzzy soft BV approach to include other precisely defined fuzzy soft sets, such as interval-valued intuitionistic fuzzy soft sets, interval type-2 fuzzy soft sets, and general type-2 fuzzy soft sets. Moreover, we plan to investigate the effectiveness of the suggested modified similarity measures across diverse domains where uncertainty is a critical consideration, encompassing areas like image processing, feature classification, decision-making, and partition clustering. A particularly noteworthy application potential for our proposed measure lies in the classification of big data, warranting in-depth exploration.

References

[1] K. T. Atanassov, *Intuitionistic Fuzzy Sets*, Springer, 1999.

[2] T. Athira, S. J. John and H. Garg, *Similarity measures of Pythagorean fuzzy soft sets and clustering analysis*, Soft Computing **27** (2023), 3007–3022.

[3] F. E. Boran and D. Akay, *A biparametric similarity measure on intuitionistic fuzzy sets with applications to pattern recognition*, Information sciences **255** (2014), 45–57.

[4] N. Çağman and D. Irfan, *Similarity measures of intuitionistic fuzzy soft sets and their decision making*, arXiv preprint arXiv:1301.0456, 2013.

[5] S.-M. Chen, *Measures of similarity between vague sets*, Fuzzy sets and Systems **74** (1995), 217–223.

[6] S.-M. Chen, M.-S. Yeh and P.-Y. Hsiao, *A comparison of similarity measures of fuzzy values*, Fuzzy sets and systems **72** (1995), 79–89.

[7] F. Ba.sar, B. Altay, M. Mursaleen, *Some generalizations of the space BVP of p-bounded variation sequences*, Nonlinear Analysis: Theory, Methods and Applications **68** (2008) 273–287.

[8] N. R. Das, S. S. Nur, A. Tiwari, Q. D. Lohani, *A novel Sugeno integral based similarity measure of generalized intuitionistic fuzzy soft sets and its application in decision making*, in: 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2022, pp. 1–7.

[9] B. Dinda and T. K. Samanta, *Relations on intuitionistic fuzzy soft sets*, arXiv preprint arXiv:1202.4649, 2012.

[10] Z. Gong and C. Wu, *Bounded variation, absolute continuity and absolute integrability for fuzzy-number-valued functions*, Fuzzy Sets and Systems **129** (2002), 83–94.

[11] K. Hirota, *The bounded variation quantity (BVQ) and its application to feature extraction*, Pattern Recognition **15** (1982), 93–101.

[12] D. H. Hong and C. Kim, *A note on similarity measures between vague sets and between elements*, Information sciences **115** (1999), 83–96.

[13] W.-L. Hung and M.-S. Yang, *Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance*, Pattern recognition letters **25** (2004), 1603–1611.

[14] W.-L. Hung and M.-S. Yang, *Similarity measures of intuitionistic fuzzy sets based on lp metric*, International Journal of Approximate Reasoning **46** (2007), 120–136.

[15] Z. Liang and P. Shi, *Similarity measures on intuitionistic fuzzy sets*, Pattern recognition letters **24** (2003), 2687–2693.

[16] Q. D. Lohani, R. Solanki and P. K. Muhuri, *A convergence theorem and an experimental study of intuitionistic fuzzy c-mean algorithm over machine learning dataset*, Applied Soft Computing **71** (2018), 1176–1188.

[17] Q. D. Lohani, R. Solanki and P. K. Muhuri, *Novel adaptive clustering algorithms based on a probabilistic similarity measure over Atanassov intuitionistic fuzzy set*, IEEE Transactions on Fuzzy Systems **26** (2018), 3715–3729.

[18] P. K. Maji, R. Biswas and A. R. Roy, *Intuitionistic fuzzy soft sets*, Journal of fuzzy mathematics **9** (2001), 677–692.

[19] P. Majumdar and S. K. Samanta, *Similarity measure of soft sets*, New mathematics and natural computation **4** (2008), 1–12.

[20] P. Majumdar and S. K. Samanta, *Generalised fuzzy soft sets*, Computers & Mathematics with Applications **59** (2010), 1425–1432.

[21] D. Molodtsov, *Soft set theory–first results*, Computers and mathematics with applications **37** (1999), 19–31.

[22] P. Muthukumar and G. S. S. Krishnan, *A similarity measure of intuitionistic fuzzy soft sets and its application in medical diagnosis*, Applied Soft Computing **41** (2016), 148–156.

[23] Y. Narukawa and V. Torra, *Non-monotonic fuzzy measures and intuitionistic fuzzy sets*, in: Modeling Decisions for Artificial Intelligence: Third International Conference, MDAI 2006, Tarragona, Spain, April 3-5, 2006, Springer, 2006, pp. 150–160.

[24] W. Pedrycz, *Fuzzy sets in pattern recognition: methodology and methods*, Pattern recognition **23** (1990), 121–146.

[25] E. Szmidt and J. Kacprzyk, *Distances between intuitionistic fuzzy sets*, Fuzzy sets and systems **114** (2000), 505–518.

[26] E. Szmidt and J. Kacprzyk, *A similarity measure for intuitionistic fuzzy sets and its application in supporting medical diagnostic reasoning*, in: Artificial Intelligence and Soft Computing-ICAISC 2004: 7th International Conference, Zakopane, Poland, June 7-11, 2004. vol. 7, Springer, 2004, pp. 388–393.

[27] Ö. Talo and F. Basşar, *On the space $bv_p(f)$ of sequences of p-bounded variation of fuzzy numbers*, Acta Mathematica Sinica, English Series **24** (2008), 1205–1212.

[28] B. C. Tripathy and A. J. Dutta, *Bounded variation double sequence space of fuzzy real numbers*, Computers and Mathematics with Applications **59** (2010), 1031–1037.

[29] H. Wang, G. Qian and X.-Q. Feng, *Predicting consumer sentiments using online sequential extreme learning machine and intuitionistic fuzzy sets*, Neural Computing and Applications **22** (2013), 479–489.

[30] W. Wang and X. Xin, *Distance measure between intuitionistic fuzzy sets*, Pattern recognition letters **26** (2005), 2063–2069.

[31] Z. Xu, *Intuitionistic fuzzy aggregation operators*, IEEE Transactions on fuzzy systems **15** (2007), 1179–1187.

[32] Z. Xu, J. Chen and J. Wu, *Clustering algorithm for intuitionistic fuzzy sets*, Information sciences **178** (2008), 3775–3790.

[33] L. Zadeh, *Fuzzy sets*, Inform Control **8** (1965), 338–353.

[34] A. Zubair, M. S. Khan and Q. M. D. Lohani, *New bounded variation based similarity measures between Atanassov intuitionistic fuzzy sets for clustering and pattern recognition*, Applied Soft Computing **85** (2019): 105529.

Manuscript received April 20, 2024
revised August 7, 2024

N. R. Das
Department of Mathematics, South Asian University, New Delhi-110068, India
E-mail address: niher381406@gmail.com

A. Tiwari
School of Computer Science and Engineering, RV University, Bengaluru, Karnataka -560059, India
E-mail address: ashutoshtiwari591@gmail.com

Q. M. D. Lohani
Department of Mathematics, South Asian University, New Delhi-110068, India
E-mail address: danishlohani@cs.sau.ac.in