





A Similarity Measure Based on Bidirectional Subsethood for Intervals

Shaily Kabir , *Student Member, IEEE*, Christian Wagner , *Senior Member, IEEE*, Timothy Craig Havens , *Senior Member, IEEE*, and Derek T. Anderson , *Senior Member, IEEE*

Abstract—With a growing number of areas leveraging interval-valued data—including in the context of modeling human uncertainty (e.g., in cybersecurity), the capacity to accurately and systematically compare intervals for reasoning and computation is increasingly important. In practice, well established set-theoretic similarity measures, such as the Jaccard and Sørensen–Dice measures, are commonly used, whereas axiomatically, a wide breadth of possible measures have been theoretically explored. This article identifies, articulates, and addresses an inherent and so far not discussed limitation of popular measures—their tendency to be subject to *aliasing*—where they return the same similarity value for very different sets of intervals. The latter risks counter-intuitive results and poor-automated reasoning in real-world applications dependent on systematically comparing interval-valued system variables or states. Given this, we introduce new axioms establishing desirable properties for robust similarity measures, followed by putting forward a novel set-theoretic similarity measure based on the concept of bidirectional subsethood, which satisfies both traditional and new axioms. The proposed measure is designed to be sensitive to the variation in the size of intervals, thus avoiding *aliasing*. This article provides a detailed theoretical exploration of the new proposed measure, and systematically demonstrates its behavior using an extensive set of synthetic and real-world data. Specifically, the measure is shown to return robust outputs that follow intuition—essential for real-world applications. For example, we show that it is bounded above and below by the Jaccard and Sørensen–Dice similarity measures (when the minimum *t-norm* is used). Finally, we show that a dissimilarity or distance measure, which satisfies the properties of a metric, can easily be derived from the proposed similarity measure.

Index Terms—Distance measure (DM), interval-valued data, similarity measure (SM), subsethood.

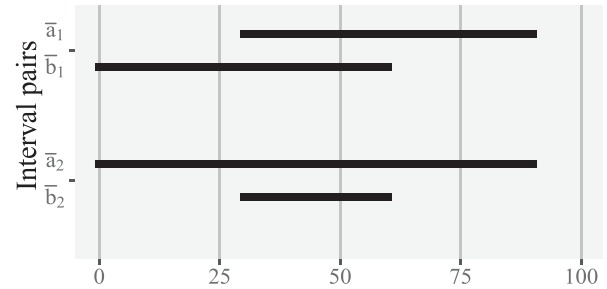


Fig. 1. Two different pairs of intervals showing *aliasing* with the Jaccard and Dice SMs.

I. INTRODUCTION

SIMILARITY measures (SMs) are widely used in many areas, including decision making, data aggregation, approximate reasoning, and machine learning. Various SMs have come into use to capture likeness among objects, though each of the measures has its own strengths and weaknesses. Similarity is commonly represented by a nonnegative real number, often between 0, meaning objects are not similar at all, and 1, meaning they are identical. SMs are generally symmetrical, though, for certain objects, similarity can be better modeled by unidirectional or asymmetric functions [1]. In addition, SMs can be transformed to capture distance using various functions [2], showing dissimilarity among objects (e.g., color images as in [3]).

Interval-valued data have recently gained much interest for the modeling of uncertainty and vagueness, particularly in the modeling of survey data [4], the representation of symbolic data [5], and the capture of natural language expressions [6], as they offer simple and efficient representation of uncertain, vague, and imprecise information. In such areas, intervals are often compared through various SMs. Among them, the Jaccard [7] and the Sørensen–Dice [8] SMs—henceforth referred to as Dice SM for convenience—are the most commonly used set-theoretic SMs in practice. These two measures provide symmetrical similarity, growing slowly from a minimum (0) to a maximum value (1) in response to an increasing degree of overlap between two closed intervals.

Nevertheless, both of these, and indeed most set-theoretic SMs frequently suffer from a so-far not discussed pitfall, best described as *aliasing*. *Aliasing* occurs where the same similarity value is generated for very different interval pairs. Fig. 1 shows an example of such interval pairs for which the Jaccard and Dice

Manuscript received March 6, 2018; revised July 13, 2018, November 11, 2018, April 28, 2019, and September 8, 2019; accepted September 12, 2019. Date of publication March 2, 2020; date of current version October 30, 2020. This work was supported in part by the UK EPSRC under Grant EP/P011918/1 Leveraging the Multi-Stakeholder Nature of Cyber Security. (Corresponding author: Shaily Kabir.)

S. Kabir and C. Wagner are with the School of Computer Science, University of Nottingham, NG7 2RD Nottingham, U.K. (e-mail: shaily.kabir@nottingham.ac.uk; christian.wagner@nottingham.ac.uk).

T. C. Havens is with the Department of Electrical and Computer Engineering and the Department of Computer Science, Michigan Technological University, Houghton, MI 49931 USA (e-mail: thavens@mtu.edu).

D. T. Anderson is with the Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211 USA (e-mail: andersondt@missouri.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TFUZZ.2019.2945249

SMs give the same similarity of 0.33 and 0.50, respectively. While SMs returning, such identical results for different sets of intervals is at least counter-intuitive, at worst it leads to incorrect inference in real-world applications. The reason why these measures exhibit *aliasing* is because their sensitivity in respect to changes in the relative size of the intervals is limited, that is, they are largely driven by the size of intersection and union. However, it is reasonable to expect that similarity would vary *both* in respect to the overlap and to the mutual similarity-in-size of the intervals.

We initially put forward the underpinnings of a new set-theoretic SM in [9] for pairs of (closed) intervals that considered their reciprocal overlapping ratios to capture similarity with a high degree of sensitivity.

The rest of this article builds on our initial work [9] while using the term subsethood rather than overlapping ratio as they are equivalent in practice. Moreover, subsethood is commonly used and well understood in the literature. The major contributions of this article are as follows.

- 1) *Aliasing* is identified and articulated as a risk and potential shortcoming affecting many popular SMs.¹
- 2) Going beyond the common axioms for being an SM, five new axioms are introduced along with their justifications to expand the axiomatic definition toward robust SMs. Furthermore, an axiomatic definition of subsethood for intervals is also provided (see Appendix).
- 3) A new set-theoretic SM is proposed that is designed to be sensitive to the variation in the size of intervals, thus avoiding *aliasing*.
- 4) The new SM is explored theoretically, showing its conformity with the expanded axiomatic definition, that is, both common and new axioms of a robust SM.
- 5) A new distance measure (DM) is derived from the new SM for estimating distance or dissimilarity between intervals, and it is proved to be a metric.
- 6) The utility of the new SM is demonstrated in the context of both synthetic and real-world cases. Using both synthetic and real-world interval-valued data, we demonstrate its intuitive behavior in line with popular SMs—specifically being bounded above and below by Jaccard and Dice—highlighting how the new SM follows expected and intuitive behavior while still addressing pitfalls of existing SMs (e.g., by avoiding *aliasing*).

The rest of this article is structured as follows. Section II briefly reviews SMs, including the Jaccard and Dice, as well as DMs. Section III discusses an axiomatic definition toward robust SMs for (closed) intervals, followed by proposing a new set-theoretic SM, which supports all axioms and then derives a new distance metric from the new SM. In Section IV, we explore the behavior of the new SM in respect to the Jaccard and Dice measures, using both synthetic and real-world interval-valued datasets. Finally, Section V concludes this article. Table I

TABLE I
ACRONYMS AND NOTATION

SM	Similarity measure
DM	Distance measure
S_J	Jaccard SM
S_D	Dice SM
S_T	Tversky's ratio model
S_h	Subsethood
S_{S_h}	Bidirectional subsethood based SM
D_{S_h}	Bidirectional subsethood based DM
X	Universe of discourse
a	Crisp set
\bar{a}	Interval $\{\bar{a} \subseteq \mathbb{R} : \bar{a} = [a^-, a^+], a^- < a^+\}$
A	Fuzzy set $\{(x, \mu_A(x)) x \in X \text{ and } \mu_A(x) \in [0, 1]\}$

presents a list of acronyms and notation used in this article to assist the reader.

II. BACKGROUND

We review SMs generally, including their most common properties arising from their axiomatic definition, followed by a detailed review of the Jaccard and Dice measures as the most widely used SMs for intervals in the literature. Next, we briefly discuss DMs and the properties required for the latter being a metric, followed by a review of subsethood, which is employed later in this article to derive a new SM. We focus throughout on measures applied to intervals in this article as the latter provide the foundation for the later extension to SMs for more complex data types, including fuzzy sets.

A. Similarity Measures

An SM $S(a, b) \rightarrow [0, 1]$ is a real-valued function that determines how two objects, a and b , are alike. Generally, the similarity between two objects is bounded by 0 and 1, where 0 means that both objects are completely different and 1 means that they are identical. The four common properties of an SM for sets a , b , and c are as follows [14].

[A1] Boundedness: $0 \leq S(a, b) \leq 1$.

[A2] Symmetry: $S(a, b) = S(b, a)$.

[A3] Reflexivity: $S(a, b) = 1 \iff a = b$.

[A4] Transitivity: If $a \subseteq b \subseteq c$ then $S(a, b) \geq S(a, c)$.

The aforementioned properties are mirrored in (and in many cases arise from) the axiomatic definitions associated with a wide variety of established SMs. Going beyond these standard and generic properties, we now provide a more detailed review of both the Jaccard and Dice SMs as the most common SMs used in practice.

1) *Jaccard SM*: The Jaccard SM [7] is one of the most widely used set-theoretic SMs, and it satisfies all of the aforementioned properties. Generally, the Jaccard similarity of two crisp sets a and b is defined as the ratio of the cardinality² of their intersection

¹Beyond both Jaccard and Dice measures and other set-theoretic SMs, such as Szymkiewicz–Simpson coefficient [10], Otsuka coefficient [11], Sokal–Sneath coefficient [12], and simple-matching coefficient [13], are also subject to *aliasing* for intervals.

²The cardinality of a set a is defined as the number of elements within a [15]. In this article, we are interchangeably using the terms “cardinality” and “size” of a set. For intervals, we also use “width.”

and the cardinality of their union

$$S_J(a, b) = \frac{|a \cap b|}{|a \cup b|}. \quad (1)$$

Using the crisp set difference operation [15], (1) can be written as

$$S_J(a, b) = \frac{|a \cap b|}{|a \cap b| + |a \setminus b| + |b \setminus a|} \quad (2)$$

where $a \setminus b$ is the set of items that are in a but not in b and $b \setminus a$ is the set of items that are in b but not in a . Equation (2) can also be derived from Tversky's parameterized ratio model of similarity [1]

$$S_T(a, b) = \frac{f(a \cap b)}{f(a \cap b) + \alpha f(a \setminus b) + \beta f(b \setminus a)}$$

by setting the nonnegative factors α and β to 1 and letting f be a cardinality function. Note that this alternative form of the Jaccard SM at (2) is relevant for showing its relationship with the Dice and other SMs, detailed in Section III. Beyond crisp sets, the Jaccard SM is used to estimate the similarity for intervals or sets of intervals such as employed for example in data fusion [16], [17] and that of fuzzy sets [18].

A closed interval \bar{a} is a set of real numbers characterized by two endpoints a^- and a^+ with $a^- < a^+$.³ The interval \bar{a} is often represented as $[a^-, a^+]$ and its cardinality, size, or width is $|\bar{a}| = |a^+ - a^-|$. For comparing the intervals \bar{a} and \bar{b} , the Jaccard SM is expressed as

$$S_J(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cup \bar{b}|} \quad (3)$$

where $|\bar{a}| \neq 0$ and $|\bar{b}| \neq 0$. $|\bar{a} \cap \bar{b}|$ is the size of the intersection between \bar{a} and \bar{b} and $|\bar{a} \cup \bar{b}|$ is the size of the interval segment(s) covering both \bar{a} and \bar{b} . Hence, $S_J(\bar{a}, \bar{b}) = 1$ when \bar{a} and \bar{b} are completely overlapping and $S_J(\bar{a}, \bar{b}) = 0$ when they are not overlapping at all. Similar to (2), we can rewrite (3) as

$$S_J(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}| + |\bar{b} \setminus \bar{a}|} \quad (4)$$

where $|\bar{a} \setminus \bar{b}|$ is the size of the interval segment of \bar{a} that is not overlapping with \bar{b} and $|\bar{b} \setminus \bar{a}|$ is the size of the interval segment of \bar{b} that is not overlapping with \bar{a} .

A fuzzy set [21] is defined as a set where the set's elements have membership ranging between 0 and 1. Formally, a type-1 fuzzy set A on a discrete and finite universe of discourse X is written as [22]

$$A = \{(x, \mu_A(x)) | x \in X\} \quad (5)$$

where $\mu_A(x) \in [0, 1]$ is the membership grade of the element x in A . For two type-1 fuzzy sets A and B on the discrete and finite universe X , the Jaccard SM can be written as [23]

$$S_J(A, B) = \frac{\sum_{i=1}^N \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^N \max(\mu_A(x_i), \mu_B(x_i))} \quad (6)$$

³Note that \bar{a} is also known as a continuous [19] or convex [20] interval.

where $\mu_A(x_i)$ and $\mu_B(x_i)$ are the membership grades of x_i in A and B , respectively. Equation (6) yields a value of 1 when the fuzzy sets are identical and 0 when they are disjoint. It is noted that the Jaccard SM has been further extended for interval-valued fuzzy sets [24], [25] and type-2 fuzzy sets [26], [27]; though, this is not discussed further here.

2) *Dice SM*: The Dice SM [8] is closely related to the Jaccard SM. To assess the similarity between two sets, it considers the ratio of the cardinality of their intersection and the average of their cardinality. Like the Jaccard similarity, it produces outputs in $[0, 1]$. Specifically, for two crisp sets a and b , the Dice similarity is expressed as

$$S_D(a, b) = \frac{|a \cap b|}{\frac{1}{2}(|a| + |b|)} \quad (7)$$

where $|a|$ is the cardinality of the set a . We can rewrite (7) by applying the crisp set difference operation [15]

$$S_D(a, b) = \frac{|a \cap b|}{|a \cap b| + \frac{1}{2}(|a \setminus b| + |b \setminus a|)}. \quad (8)$$

Note that (8) can also be obtained from Tversky's ratio model [1] when both nonnegative factors α and β are 0.5. The alternative expressions of Jaccard at (2) and Dice at (8) show clearly that the averaging operation in the denominator of (8) results in the Dice similarity always being equal to—or when sets are identical—larger than the Jaccard similarity. We expand this in Section III.

In [16] and [17], the Dice similarity is used along with the Jaccard similarity for intervals. By following (4), the Dice similarity for two intervals \bar{a} and \bar{b} can be expressed as

$$S_D(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + \frac{1}{2}(|\bar{a} \setminus \bar{b}| + |\bar{b} \setminus \bar{a}|)} \quad (9)$$

where $|\bar{a}| \neq 0$ and $|\bar{b}| \neq 0$. While less frequently used for fuzzy sets than Jaccard, the Dice SM is, for example, used in [28] and [29] for trapezoidal fuzzy numbers in the context of solving multicriteria decision-making problems.

B. Distance Measures

A DM $D(a, b) \rightarrow \mathbb{R}^+$ is a real-valued function that determines how far apart two objects a and b are. A DM is a metric when it satisfies the following properties for sets a , b , and c [30].

[B1] Nonnegativity: $D(a, b) \geq 0$.

[B2] Symmetry: $D(a, b) = D(b, a)$.

[B3] Identity: $D(a, b) = 0 \iff a = b$.

[B4] Triangle inequality: $D(a, c) \leq D(a, b) + D(b, c)$.

The Jaccard DM is complementary to the Jaccard SM and is a distance metric [31]. It is simply obtained by $D_J(a, b) = 1 - S_J(a, b)$. The Dice DM is also obtained by subtracting the Dice SM from 1, i.e., $D_D(a, b) = 1 - S_D(a, b)$. However, it is not a distance metric, but is often referred to as a semimetric as it satisfies all of the preceding properties except the triangle inequality [31]. Note that a DM is often within the range $[0, 1]$ —0 for identical objects and 1 for dissimilar objects—when it is derived from an SM bounded by $[0, 1]$.

C. Subsethood

Subsethood between two crisp sets a and b is a relation that indicates the degree to which a is a subset of b [24]. It is defined as

$$S_h(a, b) = \frac{|a \cap b|}{|a|} \quad (10)$$

where $|a \cap b|$ is the cardinality of the intersection of sets a and b , and $|a|$ is the cardinality of set a .⁴ From (10), it is clear that subsethood is smaller when more elements of set a are not part of set b and is larger when more elements of set a are part of set b . In general, (10) is bounded on the interval $[0, 1]$, where 1 means that a is a subset of b ($a \subseteq b$) and 0 means a is not a subset of b ($a \not\subseteq b$)—they are disjoint.

In a similar manner, the degree of subsethood of two closed intervals \bar{a} and \bar{b} can be defined as

$$S_h(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|} \quad (11)$$

where $|\bar{a} \cap \bar{b}|$ is the size of the intersection between \bar{a} and \bar{b} and $|\bar{a}| \neq 0$.⁵

A binary notion of subsethood for fuzzy sets is first defined in [21] where for two fuzzy sets A and B on the universe X , A is a subset of B ($A \subseteq B$) if and only if $\mu_A(x) \leq \mu_B(x) \forall x \in X$. This definition is inherently crisp— A is or is not a subset of B —which is incoherent in respect to the fuzzy set theory. Hence, many alternatives for fuzzy subsethood have been introduced. Using the set-theoretic approach, the degree to which A is a subset of B on a finite X is defined as [33]

$$S_h(A, B) = \begin{cases} 1, & A = \emptyset \\ \frac{\sum_{i=1}^n \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^n \mu_A(x_i)}, & A \neq \emptyset \end{cases} \quad (12)$$

where $\sum_{i=1}^n \min(\mu_A(x_i), \mu_B(x_i))$ is a measure of the cardinality of the intersection of membership functions of A and B , and $\sum_{i=1}^n \mu_A(x_i)$ is a measure of the cardinality of A . Contrarily, using the fuzzy implication operator \mathcal{I} , it is expressed as $S_h(A, B) = \inf_{x \in X} \mathcal{I}(\mu_A(x), \mu_B(x))$ where $\mathcal{I} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is such that $\mathcal{I}(0, 0) = \mathcal{I}(0, 1) = \mathcal{I}(1, 1) = 1$ and $\mathcal{I}(1, 0) = 0$ [34].

Furthermore, fuzzy subsethood is characterized by different axiomatizations. Initially, four axioms are proposed to define it as a binary fuzzy relation in [35]. Subsequently, Sinha and Dougherty [36] offer another set of axioms for fuzzy subsethood. We note that these axioms (except the first two axioms) are equivalent to those in [35]. However, in [37], it is argued that some fuzzy subsethoods do not maintain all axioms and new ones are proposed. For fuzzy sets $A, B, C \in X$, these axioms are as follows:

- [C1] $S_h(A, B) = 1$ if and only if $A \subseteq B$;
- [C2] if $P \leq A$, then $S_h(A, A^c) = 0$ if and only if $A = X$;
- [C3] if $A \subseteq B \subseteq C$, then $S_h(C, A) \leq S_h(B, A)$ and if $A \subseteq B$, then $S_h(C, A) \leq S_h(C, B)$;

⁴Equation (10) can also be derived from Tversky's ratio model [1] by setting $\alpha = 1$ and $\beta = 0$.

⁵It is noted that subsethood is also known as inclusion [32] and overlapping ratio [9] as it captures the overlapping ratio between intervals.

where P is the fuzzy set with $\mu_P(x) = \frac{1}{2}$ and A^c is the fuzzy set with $\mu_{A^c}(x) = 1 - \mu_A(x)$. Later, axiom [C3] is modified as “if $A \subseteq B \subseteq C$, then $S_h(C, A) \leq S_h(B, A)$ and $S_h(C, A) \leq S_h(C, B)$ ” in [38] for defining new fuzzy subsethoods and their application to cluster validity. Besides, axioms [C2] and [C3] are altered in [39] as “ $S_h(A, A^c) = 0$ if and only if $A = X$ ” and “if $A \subseteq B$, then $S_h(A, C) \geq S_h(B, C)$ and $S_h(C, A) \leq S_h(C, B)$ ” to introduce a new fuzzy DI-subsethood by aggregating implication functions, respectively.

In many cases, fuzzy subsethood axioms are extended for interval-valued, intuitionistic, and type-2 fuzzy sets. Specifically, In [40]–[42], key axioms for interval-valued fuzzy set subsethood are presented. Furthermore, in [43], an interval-valued fuzzy strong S-subsethood measure is defined by aggregating implication functions. The extension of fuzzy subsethood for intuitionistic [44] and type-2 [45] fuzzy sets are not discussed.

Remark 1: In this article, we build on the existing literature on fuzzy subsethood, particularly its key axioms to develop main properties of subsethood for intervals, see Appendix.

D. Interaction of Subsethood and SMs for Fuzzy Sets

The relationship between subsethood and SM has been explored for fuzzy sets. Particularly, Zeng and Li [46] and Li *et al.* [32] establish interchangeability between fuzzy subsethood and fuzzy set SM based on their axiomatic definitions. Zeng and Guo [41] also study similar relationship for interval-valued fuzzy sets. Furthermore, a set of axioms are proposed in [47] to define the properties of a fuzzy set SM, which is extensively used in earlier studies. Furthermore, three different SMs for fuzzy values are proposed and compared in [48] where key properties are coincided with the axioms defined in [47]. Besides, in [49] and [50], the same set of axioms to design fuzzy set SMs for comparing images with a restricted equivalence function is also considered [51].

Remark 2: A variety of works have focused on definitions and extensions of subsethood across the fuzzy set literature, including for type-1 [33], interval-valued [40], and type-2 [45] fuzzy sets. In this article, we focus on SMs for intervals rather than fuzzy sets as the latter provide the direct underpinnings for straightforward subsequent extension (via the alpha-cut decomposition representation) to fuzzy sets.

In the next section, we first introduce an expanded axiomatic definition for *robust* SMs, followed by proposing a new SM for closed intervals that fulfills said definition and provides robust comparison of intervals.

III. TOWARD ROBUST SET-THEORETIC SMs

As outlined in Section I, *aliasing* poses a so far not discussed challenge to the robust comparison of intervals. In order to address this and move toward robust SMs, we first put forward an axiomatic definition for robust SMs for (closed) intervals. Then, building on this, in Section III-B, we define a novel set-theoretic SM for intervals [9] based on their bidirectional subsethood, showing that it follows the axiomatic definition put forward. From this SM, we derive a new DM for intervals along with a

proof showing the measure is a metric. It is worth mentioning that we also define the properties of the subsethood for intervals in respect to the wider literature, for conciseness, this is provided in the Appendix.

A. Expanded Axiomatic Definition for Robust SMs on Closed Intervals

A real-valued function $S : \bar{a} \times \bar{b} \rightarrow [0, 1]$ is defined as an SM for (closed) intervals if it maintains the following axioms.

- [P1] $0 \leq S(\bar{a}, \bar{b}) \leq 1$ (boundedness).
- [P2] $S(\bar{a}, \bar{b}) = S(\bar{b}, \bar{a})$ (symmetry).
- [P3] $S(\bar{a}, \bar{b}) = 1$ if and only if $\bar{a} = \bar{b}$ (reflexivity).
- [P4] If $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$, then $S(\bar{a}, \bar{b}) \geq S(\bar{a}, \bar{c})$ and $S(\bar{b}, \bar{c}) \geq S(\bar{b}, \bar{a})$ (transitivity).
- [P5] $S(\bar{a}, \bar{b}) = 0$ if and only if $|\bar{a} \cap \bar{b}| = 0$ (disjointness).
- [P6] For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ where $|\bar{a}_1 \cap \bar{b}_1| > 0$, $|\bar{a}_2 \cap \bar{b}_2| > 0$, $|\bar{a}_1 \cup \bar{b}_1| = |\bar{a}_2 \cup \bar{b}_2|$, and $|\bar{a}_1 \cap \bar{b}_1| = |\bar{a}_2 \cap \bar{b}_2|$, $S(\bar{a}_1, \bar{b}_1) \neq S(\bar{a}_2, \bar{b}_2)$ if $|\bar{a}_1| \neq |\bar{a}_2|$ and $|\bar{b}_1| \neq |\bar{b}_2|$, and $|\bar{b}_1| \neq |\bar{a}_2|$ and $|\bar{b}_1| \neq |\bar{b}_2|$ (nonaliasing).
- [P7] $S(\bar{a}, \bar{b}) < 1$ if $\bar{a} \subset \bar{b}$ (subsethood-limit).
- [P8] For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ with $|\bar{a}_1 \cap \bar{b}_1| > 0$ and $|\bar{a}_2 \cap \bar{b}_2| > 0$, $S(\bar{a}_1, \bar{b}_1) = S(\bar{a}_2, \bar{b}_2)$ if $\bar{a}_2 = n \times \bar{a}_1 = [n \times a_1^-, n \times a_1^+]$ and $\bar{b}_2 = n \times \bar{b}_1 = [n \times b_1^-, n \times b_1^+]$ where $n > 0$ is a scaling factor (scaling-invariance).
- [P9] For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ where $|\bar{a}_1| = |\bar{b}_1| = |\bar{a}_2| = |\bar{b}_2|$, $S(\bar{a}_1, \bar{b}_1) < S(\bar{a}_2, \bar{b}_2)$ if $|\bar{a}_1 \cap \bar{b}_1| < |\bar{a}_2 \cap \bar{b}_2|$ (increased-overlap).

1) *Justification of the Axioms:* The axioms [P1]–[P4] are derived following the common axioms [A1]–[A4] stated in Section II-A. Axiom [P5] is put forward for disjoint intervals where their similarity should be 0 as they are not overlapping. The *aliasing* example discussed in Section I calls for the axiom [P6]. Axiom [P7] is presented for intervals when one interval is a proper subset of the other. Here, the similarity should be less than 1 as both intervals are not equal. Axiom [P8] is introduced for intervals where they are scaled up by a factor. In this case, the ratio of overlapping and nonoverlapping segments stays the same while scaling; thus, the similarity between intervals should remain constant. Finally, axiom [P9] suggests higher similarity for equal-sized intervals when their overlap increases.

B. SM Based on Bidirectional Subsethood

As noted in Section I, the motivation behind the new set-theoretic SM is to establish an SM that is sensitive to potentially (very) different cardinalities of the sets being compared, in particular, when one is a subset of the other. Thus, the proposed SM takes into consideration the reciprocal subsethood (i.e., the overlapping ratio) of both sets/intervals being compared, in order to estimate their overall similarity.

Definition 1: The bidirectional subsethood based SM S_{S_h} for a pair of intervals \bar{a} and \bar{b} is the t -norm of their reciprocal subsethoods $S_h(\bar{a}, \bar{b})$ and $S_h(\bar{b}, \bar{a})$, i.e.,

$$S_{S_h}(\bar{a}, \bar{b}) = \star(S_h(\bar{a}, \bar{b}), S_h(\bar{b}, \bar{a})) \quad (13)$$

where \star is a t -norm.

In this article, we use \star to refer to either minimum t -norm or product t -norm, whereas using “ \wedge ” and “ Π ” to indicate the minimum and product t -norms, respectively. In future, we will explore other t -norms beyond minimum and product.

Theorem 1: Consider a minimum (\wedge) or product (Π) t -norm \star and the subsethood for intervals S_h . Then, $S_{S_h}(\bar{a}, \bar{b}) = \star(S_h(\bar{a}, \bar{b}), S_h(\bar{b}, \bar{a}))$ is an SM for intervals \bar{a} and \bar{b} satisfying all axioms [P1]–[P9].

Proof: The proofs are given per axioms in Section III-C. ■ Similar to (4) and (9), we can rewrite (13) as

$$S_{S_h}(\bar{a}, \bar{b}) = \star \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|} \right) \quad (14)$$

where $|\bar{a} \setminus \bar{b}|$ is the size of nonoverlapping segment(s) of \bar{a} as to \bar{b} and vice versa for $|\bar{b} \setminus \bar{a}|$. Also, $|\bar{a}| \neq 0$ and $|\bar{b}| \neq 0$.

C. Properties of the Proposed SM

We now explore the properties of the proposed bidirectional subsethood based SM $S_{S_h}(\bar{a}, \bar{b})$.⁶

Theorem 2 (Boundedness): $0 \leq S_{S_h}(\bar{a}, \bar{b}) \leq 1$.

Proof: All t -norms are bounded by 0 and 1 [52]. This is also true for the subsethood S_h (see Appendix). This eventually means that $S_{S_h}(\bar{a}, \bar{b})$ is always within [0, 1], thus addressing the axiom [P1]. ■

Theorem 3 (Symmetry): $S_{S_h}(\bar{a}, \bar{b}) = S_{S_h}(\bar{b}, \bar{a})$.

Proof: Any t -norm (\star) is symmetric [52]. Therefore, $S_{S_h}(\bar{a}, \bar{b})$ is also symmetric. The S_{S_h} SM, thus, satisfies the axiom [P2]. ■

Theorem 4 (Reflexivity): $S_{S_h}(\bar{a}, \bar{b}) = 1 \iff \bar{a} = \bar{b}$.

Proof: If $\bar{a} = \bar{b}$, then $S_h(\bar{a}, \bar{b}) = S_h(\bar{b}, \bar{a}) = 1$. From the boundary conditions of the t -norm (\star) [52], $\star(1, 1) = 1$, thus making $S_{S_h}(\bar{a}, \bar{b}) = 1$. Alternatively, $S_{S_h}(\bar{a}, \bar{b}) = 1$ means that both $S_h(\bar{a}, \bar{b})$ and $S_h(\bar{b}, \bar{a})$ are equal to 1. This only happens when \bar{a} and \bar{b} are identical intervals. Here, the S_{S_h} SM maintains the axiom [P3]. ■

Theorem 5 (Transitivity): If $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$, then $S_{S_h}(\bar{a}, \bar{b}) \geq S_{S_h}(\bar{a}, \bar{c})$ and $S_{S_h}(\bar{b}, \bar{c}) \geq S_{S_h}(\bar{a}, \bar{c})$.

Proof: if $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$, then

$$S_{S_h}(\bar{a}, \bar{b}) = \star \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{b}|} \right) = \star \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{b}|} \right) = \frac{|\bar{a}|}{|\bar{b}|}$$

$$S_{S_h}(\bar{a}, \bar{c}) = \star \left(\frac{|\bar{a} \cap \bar{c}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{c}|}{|\bar{c}|} \right) = \star \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{c}|} \right) = \frac{|\bar{a}|}{|\bar{c}|}.$$

As $\bar{b} \subseteq \bar{c}$, it follows that $|\bar{b}| \leq |\bar{c}|$. Therefore, $\frac{|\bar{a}|}{|\bar{b}|} \geq \frac{|\bar{a}|}{|\bar{c}|}$, and hence, $S_{S_h}(\bar{a}, \bar{b}) \geq S_{S_h}(\bar{a}, \bar{c})$. Again

$$S_{S_h}(\bar{b}, \bar{c}) = \star \left(\frac{|\bar{b} \cap \bar{c}|}{|\bar{b}|}, \frac{|\bar{b} \cap \bar{c}|}{|\bar{c}|} \right) = \star \left(\frac{|\bar{b}|}{|\bar{b}|}, \frac{|\bar{b}|}{|\bar{c}|} \right) = \frac{|\bar{b}|}{|\bar{c}|}.$$

⁶Note that all properties discussed in this article are maintained by the proposed S_{S_h} SM with both minimum (\wedge) and product (Π) t -norms, unless otherwise specified.

As $\bar{a} \subseteq \bar{b}$, it follows that $|\bar{b}| \geq |\bar{a}|$. Hence, $\frac{|\bar{b}|}{|\bar{c}|} \geq \frac{|\bar{a}|}{|\bar{c}|}$ implying that $S_{S_h}(\bar{b}, \bar{c}) \geq S_{S_h}(\bar{a}, \bar{c})$. In this case, the new SM satisfies the axiom [P4]. ■

Theorem 6 (Disjointness): $S_{S_h}(\bar{a}, \bar{b}) = 0$, if and only if \bar{a} and \bar{b} are disjoint.

Proof: If \bar{a} and \bar{b} are disjoint, then $|\bar{a} \cap \bar{b}| = 0$. Hence, $S_h(\bar{a}, \bar{b}) = 0$ and $S_h(\bar{b}, \bar{a}) = 0$, implying that $S_{S_h}(\bar{a}, \bar{b}) = \star(0, 0) = 0$. Again, $S_{S_h}(\bar{a}, \bar{b}) = 0$ only when both $S_h(\bar{a}, \bar{b})$ and $S_h(\bar{b}, \bar{a})$ are equal to 0. This only happens when \bar{a} and \bar{b} are not overlapping. Here, the proof of this theorem shows that the new SM follows the axiom [P5]. ■

Theorem 7 (Nonaliasing): For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ where $|\bar{a}_1 \cap \bar{b}_1| > 0$, $|\bar{a}_2 \cap \bar{b}_2| > 0$, $|\bar{a}_1 \cup \bar{b}_1| = |\bar{a}_2 \cup \bar{b}_2|$, and $|\bar{a}_1 \cap \bar{b}_1| = |\bar{a}_2 \cap \bar{b}_2|$, $S_{S_h}(\bar{a}_1, \bar{b}_1) \neq S_{S_h}(\bar{a}_2, \bar{b}_2)$ if $|\bar{a}_1| \neq |\bar{a}_2|$ and $|\bar{a}_1| \neq |\bar{b}_2|$, and $|\bar{b}_1| \neq |\bar{a}_2|$ and $|\bar{b}_1| \neq |\bar{b}_2|$.

Proof: For the pair $\{\bar{a}_1, \bar{b}_1\}$, $S_h(\bar{a}_1, \bar{b}_1) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}$ and $S_h(\bar{b}_1, \bar{a}_1) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}$. Thus

$$S_{S_h}(\bar{a}_1, \bar{b}_1) = \star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \right).$$

Again, for the pair $\{\bar{a}_2, \bar{b}_2\}$, $S_h(\bar{a}_2, \bar{b}_2) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}$ and $S_h(\bar{b}_2, \bar{a}_2) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}$. Thus

$$S_{S_h}(\bar{a}_2, \bar{b}_2) = \star \left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|} \right).$$

As $|\bar{a}_1 \cap \bar{b}_1| = |\bar{a}_2 \cap \bar{b}_2|$, therefore

$$S_{S_h}(\bar{a}_2, \bar{b}_2) \equiv \star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|} \right).$$

Given that $|\bar{a}_1| \neq |\bar{a}_2|$ and $|\bar{a}_1| \neq |\bar{b}_2|$, it implies that

$$\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|} \text{ and } \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|}.$$

Similarly, $|\bar{b}_1| \neq |\bar{a}_2|$ and $|\bar{b}_1| \neq |\bar{b}_2|$ meaning that

$$\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|} \text{ and } \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \neq \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|}.$$

Therefore

$$\star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \right) \neq \star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_2|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_2|} \right).$$

Hence, $S_{S_h}(\bar{a}_1, \bar{b}_1) \neq S_{S_h}(\bar{a}_2, \bar{b}_2)$, thus satisfying the axiom [P6]. ■

Theorem 8 (Subsethood-limit): $S_{S_h}(\bar{a}, \bar{b}) < 1$ if $\bar{a} \subset \bar{b}$.

Proof: If $\bar{a} \subset \bar{b}$, then $|\bar{a}| < |\bar{b}|$ and $|\bar{a} \cap \bar{b}| = |\bar{a}|$. Thus, $S_h(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|} = \frac{|\bar{a}|}{|\bar{a}|} = 1$ and $S_h(\bar{b}, \bar{a}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{b}|} = \frac{|\bar{a}|}{|\bar{b}|}$. It means that $S_{S_h}(\bar{a}, \bar{b}) = \star(1, \frac{|\bar{a}|}{|\bar{b}|}) = \frac{|\bar{a}|}{|\bar{b}|}$. As $|\bar{a}| < |\bar{b}|$, then $\frac{|\bar{a}|}{|\bar{b}|} < 1$. Hence, $S_{S_h}(\bar{a}, \bar{b}) < 1$. Here, the new SM, thus, addresses the axiom [P7]. ■

Theorem 9 (Scaling-invariance): For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ with $|\bar{a}_1 \cap \bar{b}_1| > 0$ and $|\bar{a}_2 \cap \bar{b}_2| > 0$, $S_{S_h}(\bar{a}_1, \bar{b}_1) = S_{S_h}(\bar{a}_2, \bar{b}_2)$ if $\bar{a}_2 = n \times \bar{a}_1 = [n \times a_1^-, n \times a_1^+]$ and $\bar{b}_2 = n \times \bar{b}_1 = [n \times b_1^-, n \times b_1^+]$ where $n > 0$ is a scaling factor.

Proof: For the pair $\{\bar{a}_1, \bar{b}_1\}$, $S_h(\bar{a}_1, \bar{b}_1) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}$ and $S_h(\bar{b}_1, \bar{a}_1) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}$. Thus

$$S_{S_h}(\bar{a}_1, \bar{b}_1) = \star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \right).$$

Again, for the pair $\{\bar{a}_2, \bar{b}_2\}$, $S_h(\bar{a}_2, \bar{b}_2) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}$, and $S_h(\bar{b}_2, \bar{a}_2) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}$. Thus

$$S_{S_h}(\bar{a}_2, \bar{b}_2) = \star \left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|} \right).$$

Given that $\bar{a}_2 = n \times \bar{a}_1$ and $\bar{b}_2 = n \times \bar{b}_1$. It implies that $|\bar{a}_2 \cap \bar{b}_2| = |(n \times \bar{a}_1) \cap (n \times \bar{b}_1)| = n \times |\bar{a}_1 \cap \bar{b}_1|$. Therefore

$$\begin{aligned} S_{S_h}(\bar{a}_2, \bar{b}_2) &= \star \left(\frac{n \times |\bar{a}_1 \cap \bar{b}_1|}{n \times |\bar{a}_1|}, \frac{n \times |\bar{a}_1 \cap \bar{b}_1|}{n \times |\bar{b}_1|} \right) \\ &= \star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \right) = S_{S_h}(\bar{a}_1, \bar{b}_1). \end{aligned}$$

Hence, $S_{S_h}(\bar{a}_1, \bar{b}_1) = S_{S_h}(\bar{a}_2, \bar{b}_2)$. Here, the new SM follows the axiom [P8]. ■

Theorem 10 (Increased-overlap): For two interval pairs $\{\bar{a}_1, \bar{b}_1\}$ and $\{\bar{a}_2, \bar{b}_2\}$ where $|\bar{a}_1| = |\bar{b}_1| = |\bar{a}_2| = |\bar{b}_2|$, $S_{S_h}(\bar{a}_1, \bar{b}_1) < S_{S_h}(\bar{a}_2, \bar{b}_2)$ if $|\bar{a}_1 \cap \bar{b}_1| < |\bar{a}_2 \cap \bar{b}_2|$.

Proof: For the pair $\{\bar{a}_1, \bar{b}_1\}$, $S_h(\bar{a}_1, \bar{b}_1) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}$, and $S_h(\bar{b}_1, \bar{a}_1) = \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|}$. Thus

$$S_{S_h}(\bar{a}_1, \bar{b}_1) = \star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \right).$$

For the pair $\{\bar{a}_2, \bar{b}_2\}$, $S_h(\bar{a}_2, \bar{b}_2) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}$ and $S_h(\bar{b}_2, \bar{a}_2) = \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|}$. Thus

$$S_{S_h}(\bar{a}_2, \bar{b}_2) = \star \left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_2|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_2|} \right).$$

As $|\bar{a}_1| = |\bar{b}_1| = |\bar{a}_2| = |\bar{b}_2|$, therefore

$$S_{S_h}(\bar{a}_2, \bar{b}_2) \equiv \star \left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_1|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_1|} \right).$$

Given that $|\bar{a}_1 \cap \bar{b}_1| < |\bar{a}_2 \cap \bar{b}_2|$, it implies that $\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|} < \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_1|}$ and $\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} < \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_1|}$. Therefore

$$\star \left(\frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{a}_1|}, \frac{|\bar{a}_1 \cap \bar{b}_1|}{|\bar{b}_1|} \right) < \star \left(\frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{a}_1|}, \frac{|\bar{a}_2 \cap \bar{b}_2|}{|\bar{b}_1|} \right).$$

Hence, $S_{S_h}(\bar{a}_1, \bar{b}_1) < S_{S_h}(\bar{a}_2, \bar{b}_2)$. Here, the S_{S_h} SM addresses the axiom [P9]. ■

Theorem 11: $S_{S_h}(\bar{a}, \bar{b})$ is bounded by the Jaccard and Dice SMs when \star is the minimum (\wedge) t -norm. That is, $S_J(\bar{a}, \bar{b}) \leq S_{S_h \wedge}(\bar{a}, \bar{b}) \leq S_D(\bar{a}, \bar{b})$.

Proof: For the interval pair $\{\bar{a}, \bar{b}\}$, consider the formulations of the SMs at (4), (9), and (14).

To prove this theorem, we consider four cases: $\bar{a} = \bar{b}$, $\bar{a} \cap \bar{b} = \emptyset$, $\bar{a} \subset \bar{b}$, and $\bar{a} \cap \bar{b} \neq \emptyset$, and $\bar{a} \not\subset \bar{b}$ and $\bar{b} \not\subset \bar{a}$.

Case 1: If $\bar{a} = \bar{b}$, then all three measures yield a similarity of 1. That is, $S_J(\bar{a}, \bar{b}) = S_D(\bar{a}, \bar{b}) = S_{S_h \wedge}(\bar{a}, \bar{b}) = 1$.

Case 2: If $\bar{a} \cap \bar{b} = \emptyset$, all three measures result in a similarity of 0. Thus, $S_J(\bar{a}, \bar{b}) = S_D(\bar{a}, \bar{b}) = S_{S_h \wedge}(\bar{a}, \bar{b}) = 0$.

Case 3: If $\bar{a} \subset \bar{b}$ (subset), then $|\bar{a} \cap \bar{b}| = |\bar{a}|$. With respect to \bar{b} , there is no nonoverlap segment of \bar{a} ; hence, $|\bar{a} \setminus \bar{b}| = 0$. Inversely, there is a nonoverlap segment of \bar{b} as to \bar{a} ; thus, $|\bar{b} \setminus \bar{a}| \neq 0$. In this case, the three SMs can be simplified to

$$S_J(\bar{a}, \bar{b}) = \frac{|\bar{a}|}{|\bar{a}| + |\bar{b} \setminus \bar{a}|}$$

$$S_D(\bar{a}, \bar{b}) = \frac{|\bar{a}|}{|\bar{a}| + \frac{1}{2} |\bar{b} \setminus \bar{a}|}$$

$$S_{S_h \wedge}(\bar{a}, \bar{b}) = \wedge \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{a}| + |\bar{b} \setminus \bar{a}|} \right) = \frac{|\bar{a}|}{|\bar{a}| + |\bar{b} \setminus \bar{a}|}$$

which implies that

$$S_J(\bar{a}, \bar{b}) = S_{S_h \wedge}(\bar{a}, \bar{b}) < S_D(\bar{a}, \bar{b}).$$

Case 4: If $\bar{a} \cap \bar{b} \neq \emptyset$, and $\bar{a} \not\subset \bar{b}$ and $\bar{b} \not\subset \bar{a}$ (\bar{a} and \bar{b} are partially overlapping), then assume the case $|\bar{a}| \leq |\bar{b}|$. It implies that $|\bar{a}| - |\bar{a} \cap \bar{b}| \leq |\bar{b}| - |\bar{a} \cap \bar{b}| \Rightarrow |\bar{a} \setminus \bar{b}| \leq |\bar{b} \setminus \bar{a}|$. The three SMs are

$$S_J(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}| + |\bar{b} \setminus \bar{a}|}$$

$$S_D(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + \frac{1}{2} (|\bar{a} \setminus \bar{b}| + |\bar{b} \setminus \bar{a}|)}$$

$$\begin{aligned} S_{S_h \wedge}(\bar{a}, \bar{b}) &= \wedge \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|} \right) \\ &= \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|}, \text{ as } |\bar{a} \setminus \bar{b}| \leq |\bar{b} \setminus \bar{a}|. \end{aligned}$$

It is true that

$$\frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}| + |\bar{b} \setminus \bar{a}|} < \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + \frac{1}{2} (|\bar{a} \setminus \bar{b}| + |\bar{b} \setminus \bar{a}|)}$$

thus $S_J(\bar{a}, \bar{b}) < S_D(\bar{a}, \bar{b})$. Again, it is clear that

$$\frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}| + |\bar{b} \setminus \bar{a}|} < \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|}$$

implying that $S_J(\bar{a}, \bar{b}) < S_{S_h \wedge}(\bar{a}, \bar{b})$. Also

$$\begin{aligned} \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|} &= \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + \frac{1}{2} |\bar{b} \setminus \bar{a}| + \frac{1}{2} |\bar{b} \setminus \bar{a}|} \\ &\leq \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + \frac{1}{2} |\bar{a} \setminus \bar{b}| + \frac{1}{2} |\bar{b} \setminus \bar{a}|}, \text{ as } |\bar{a} \setminus \bar{b}| \leq |\bar{b} \setminus \bar{a}| \end{aligned}$$

indicating that $S_{S_h \wedge}(\bar{a}, \bar{b}) \leq S_D(\bar{a}, \bar{b})$. Hence, $S_J(\bar{a}, \bar{b}) < S_{S_h \wedge}(\bar{a}, \bar{b}) \leq S_D(\bar{a}, \bar{b})$. Note that for the case $|\bar{b}| \leq |\bar{a}|$, the same procedure can be used to prove the aforementioned relation. ■

D. DM Based on Bidirectional Subsethood

A new DM $D_{S_h}(\bar{a}, \bar{b})$ can easily be derived from the S_{S_h} measure at (13) by taking its complement, capturing the dissimilarity between both intervals

$$D_{S_h}(\bar{a}, \bar{b}) = 1 - S_{S_h}(\bar{a}, \bar{b}). \quad (15)$$

Alternatively, (15) can be written as

$$\begin{aligned} D_{S_h}(\bar{a}, \bar{b}) &= 1 - \star \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|} \right) \\ &= \star \left(1 - \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}|}, 1 - \frac{|\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|} \right) \\ &= \star \left(\frac{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}| - |\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}|}, \frac{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}| - |\bar{a} \cap \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|} \right) \\ &= \star \left(\frac{|\bar{a} \setminus \bar{b}|}{|\bar{a} \cap \bar{b}| + |\bar{a} \setminus \bar{b}|}, \frac{|\bar{b} \setminus \bar{a}|}{|\bar{a} \cap \bar{b}| + |\bar{b} \setminus \bar{a}|} \right). \end{aligned}$$

Note that this alternative form of the proposed $D_{S_h}(\bar{a}, \bar{b})$ measure can now directly be used in pattern recognition problems with sets/intervals, such as classification and clustering.

We next discuss the essential properties of the $D_{S_h}(\bar{a}, \bar{b})$ measure in terms of it being a metric.

Theorem 12: $D_{S_h}(\bar{a}, \bar{b}) = 1 - S_{S_h}(\bar{a}, \bar{b})$ is a metric.

Proof: To prove that the DM, $D_{S_h}(\bar{a}, \bar{b})$ is a metric, we need to show that it satisfies the following properties for the intervals \bar{a} , \bar{b} , and \bar{c} where \star is the minimum (\wedge) or product (Π) t -norm.

- $D_{S_h}(\bar{a}, \bar{b}) \geq 0$ (nonnegativity).
- $D_{S_h}(\bar{a}, \bar{b}) = D_{S_h}(\bar{b}, \bar{a})$ (symmetry).
- $D_{S_h}(\bar{a}, \bar{b}) = 0 \iff \bar{a} = \bar{b}$ (identity).
- $D_{S_h}(\bar{a}, \bar{c}) \leq D_{S_h}(\bar{a}, \bar{b}) + D_{S_h}(\bar{b}, \bar{c})$ (triangle inequality).

We provide proofs for all the aforementioned properties (a)–(d) in the following.

(a) By Theorem 3, $S_{S_h}(\bar{a}, \bar{b})$ is always bounded by $[0, 1]$. As $D_{S_h}(\bar{a}, \bar{b}) = 1 - S_{S_h}(\bar{a}, \bar{b})$, it implies that $0 \leq D_{S_h}(\bar{a}, \bar{b}) \leq 1$, thus meeting the nonnegativity property.

(b) By Theorem 4, $S_{S_h}(\bar{a}, \bar{b})$ is always symmetric; thus, $D_{S_h}(\bar{a}, \bar{b})$ is also symmetric.

(c) Theorem 5 shows that if $\bar{a} = \bar{b}$, then $S_{S_h}(\bar{a}, \bar{b})$ is always 1. Hence, $1 - S_{S_h}(\bar{a}, \bar{b})$ makes $D_{S_h}(\bar{a}, \bar{b}) = 0$. Alternatively, If $\bar{a} \neq \bar{b}$, then $S_{S_h}(\bar{a}, \bar{b}) \neq 1$, meaning $D_{S_h}(\bar{a}, \bar{b}) \neq 0$.

(d) For the intervals \bar{a} , \bar{b} , and \bar{c} , we consider $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$. It implies that $|\bar{a}| \leq |\bar{b}| \leq |\bar{c}|$. To prove this theorem, we apply the formulation of the S_{S_h} measure at (14).

Case 1: When all three intervals are equal ($\bar{a} = \bar{b} = \bar{c}$), $S_{S_h}(\bar{a}, \bar{b}) = S_{S_h}(\bar{b}, \bar{c}) = S_{S_h}(\bar{a}, \bar{c}) = 1$. Accordingly, $D_{S_h}(\bar{a}, \bar{b}) = D_{S_h}(\bar{b}, \bar{c}) = D_{S_h}(\bar{a}, \bar{c}) = 0$, thus satisfying the triangle inequality.

Case 2: When $\bar{a} \subset \bar{b} \subset \bar{c}$, it implies that $|\bar{a}| < |\bar{b}| < |\bar{c}|$. Now, we can calculate the similarity and distance for each pair of intervals by applying (14).

(i) For $\bar{a} \subset \bar{b}$

$$\begin{aligned} S_{S_h}(\bar{a}, \bar{b}) &= \star \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{b}|} \right) \\ &= \star \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{b}|} \right) = \frac{|\bar{a}|}{|\bar{b}|}. \end{aligned}$$

$$\text{Thus, } D_{S_h}(\bar{a}, \bar{b}) = 1 - S_{S_h}(\bar{a}, \bar{b}) = 1 - \frac{|\bar{a}|}{|\bar{b}|}.$$

(ii) For $\bar{b} \subset \bar{c}$

$$S_{S_h}(\bar{b}, \bar{c}) = \frac{|\bar{b}|}{|\bar{c}|}.$$

$$\text{Thus, } D_{S_h}(\bar{b}, \bar{c}) = 1 - S_{S_h}(\bar{b}, \bar{c}) = 1 - \frac{|\bar{b}|}{|\bar{c}|}.$$

(iii) For $\bar{a} \subset \bar{c}$ (by the transitive property of subsets [15])

$$S_{S_h}(\bar{a}, \bar{c}) = \frac{|\bar{a}|}{|\bar{c}|}.$$

$$\text{Thus, } D_{S_h}(\bar{a}, \bar{c}) = 1 - S_{S_h}(\bar{a}, \bar{c}) = 1 - \frac{|\bar{a}|}{|\bar{c}|}.$$

Now, we have to show that

$$D_{S_h}(\bar{a}, \bar{c}) < D_{S_h}(\bar{a}, \bar{b}) + D_{S_h}(\bar{b}, \bar{c}).$$

By placing the distance of each pair of intervals in the aforementioned equation, we get the following:

$$\begin{aligned} 1 - \frac{|\bar{a}|}{|\bar{c}|} &< 1 - \frac{|\bar{a}|}{|\bar{b}|} + 1 - \frac{|\bar{b}|}{|\bar{c}|} \\ \iff 1 - \frac{|\bar{a}|}{|\bar{c}|} - 1 + \frac{|\bar{b}|}{|\bar{c}|} &< 1 - \frac{|\bar{a}|}{|\bar{b}|} \\ \iff -\frac{|\bar{a}|}{|\bar{c}|} + \frac{|\bar{b}|}{|\bar{c}|} &< 1 - \frac{|\bar{a}|}{|\bar{b}|} \\ \iff \frac{|\bar{b}| - |\bar{a}|}{|\bar{c}|} &< \frac{|\bar{b}| - |\bar{a}|}{|\bar{b}|} \\ \iff \frac{1}{|\bar{c}|} &< \frac{1}{|\bar{b}|}. \end{aligned}$$

It is true as $|\bar{b}| < |\bar{c}|$, thereby satisfying the triangle inequality. Thus, $D_{S_h}(\bar{a}, \bar{c}) \leq D_{S_h}(\bar{a}, \bar{b}) + D_{S_h}(\bar{b}, \bar{c})$. ■

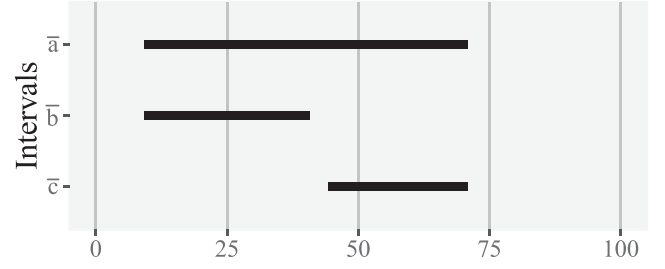


Fig. 2. Example that shows to satisfy the triangle inequality by the D_{S_h} measure.

TABLE II
SIMILARITY AND DISTANCE BETWEEN THE INTERVALS OF FIG. 2

Interval pair	$S_{S_h \wedge}$	$S_{S_h \Pi}$	$D_{S_h \wedge}$	$D_{S_h \Pi}$
$\{\bar{a}, \bar{b}\}$	0.50	0.50	0.50	0.50
$\{\bar{a}, \bar{c}\}$	0.42	0.42	0.58	0.58
$\{\bar{b}, \bar{c}\}$	0.0	0.0	1.0	1.0

Example satisfying triangle inequality: Consider three intervals \bar{a} , \bar{b} , and \bar{c} in Fig. 2 where $\bar{b}, \bar{c} \subset \bar{a}$ and $\bar{b} \cap \bar{c} = \emptyset$. Table II presents the similarity and distance between each pair of the three intervals using the S_{S_h} and D_{S_h} measures, respectively, with the minimum (\wedge) and product (Π) t -norms. Note that the same similarity and distance results are received using both t -norms. From the distance results, it is observed that for the interval pair $\{\bar{a}, \bar{c}\}$, their distance—which is 0.58—is less than the summed distances of the pairs $\{\bar{a}, \bar{b}\}$ and $\{\bar{b}, \bar{c}\}$ —which is 1.50—implying $D_{S_h}(\bar{a}, \bar{c}) < D_{S_h}(\bar{a}, \bar{b}) + D_{S_h}(\bar{b}, \bar{c})$. This relation is also maintained for the interval pairs $\{\bar{a}, \bar{b}\}$ and $\{\bar{b}, \bar{c}\}$, thereby demonstrating the D_{S_h} measure meeting the triangle inequality.

Theorem 13: $D_{S_h}(\bar{a}, \bar{b})$ follows the property of transitivity. That is, $D_{S_h}(\bar{a}, \bar{b}) \leq D_{S_h}(\bar{a}, \bar{c})$ when $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$.

Proof: When $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$, then

$$S_{S_h}(\bar{a}, \bar{b}) = \star \left(\frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{b}|}{|\bar{b}|} \right) = \star \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{b}|} \right) = \frac{|\bar{a}|}{|\bar{b}|}$$

$$S_{S_h}(\bar{a}, \bar{c}) = \star \left(\frac{|\bar{a} \cap \bar{c}|}{|\bar{a}|}, \frac{|\bar{a} \cap \bar{c}|}{|\bar{c}|} \right) = \star \left(\frac{|\bar{a}|}{|\bar{a}|}, \frac{|\bar{a}|}{|\bar{c}|} \right) = \frac{|\bar{a}|}{|\bar{c}|}.$$

Thus

$$D_{S_h}(\bar{a}, \bar{b}) = 1 - S_{S_h}(\bar{a}, \bar{b}) = 1 - \frac{|\bar{a}|}{|\bar{b}|}$$

$$D_{S_h}(\bar{a}, \bar{c}) = 1 - S_{S_h}(\bar{a}, \bar{c}) = 1 - \frac{|\bar{a}|}{|\bar{c}|}.$$

As $\bar{b} \subseteq \bar{c}$, it follows that $|\bar{b}| \leq |\bar{c}|$. Therefore, $\frac{|\bar{a}|}{|\bar{b}|} \geq \frac{|\bar{a}|}{|\bar{c}|}$, which implies that

$$1 - \frac{|\bar{a}|}{|\bar{b}|} \leq 1 - \frac{|\bar{a}|}{|\bar{c}|}.$$

Hence $D_{S_h}(\bar{a}, \bar{b}) \leq D_{S_h}(\bar{a}, \bar{c})$. ■

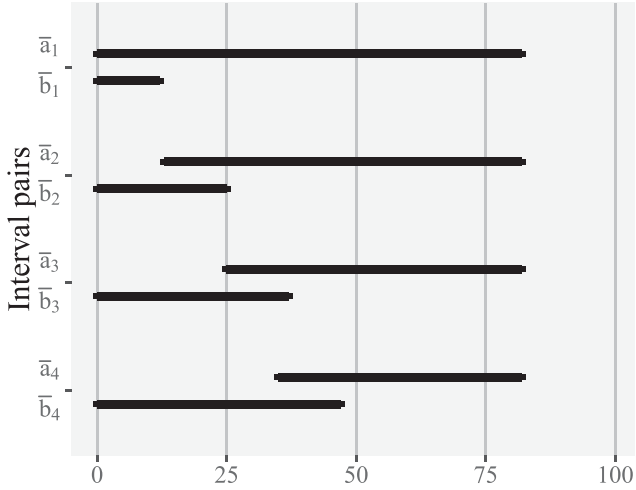


Fig. 3. Interval pairs used to demonstrate the *aliasing* of SMs.

TABLE III
 S_J AND S_D MEASURES EXHIBIT ALIASING FOR THE INTERVAL PAIRS IN FIG. 3

Interval pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$
$\{\bar{a}_1, \bar{b}_1\}$	0.15	0.26	0.15	0.15
$\{\bar{a}_2, \bar{b}_2\}$	0.15	0.26	0.17	0.08
$\{\bar{a}_3, \bar{b}_3\}$	0.15	0.26	0.21	0.07
$\{\bar{a}_4, \bar{b}_4\}$	0.15	0.26	0.26	0.06

IV. DEMONSTRATION AND ANALYSIS

We now demonstrate and analyze the behavior of the proposed SM S_{S_h} with the minimum (\wedge) and product (Π) t -norms in the context of the S_J and S_D SMs for a set of synthetic examples in the first part and with a real dataset in the last part. Here, we alter different features of interval pairs to explore how well these three measures perform or follow intuitive results. In particular, we focus on the following aspects.

- 1) Propensity to exhibiting *aliasing* in response to variations in interval sizes.
- 2) Behavior in respect to intervals where one is a complete subset of the other.
- 3) Behavior in respect to intervals of equal sizes and overlapping ratio.
- 4) Response to variations in interval size while maintaining the same level of subsethood.
- 5) Response to linear increase in the overlap of intervals.

A. Synthetic Dataset Based Demonstration

For each of the aforementioned cases, a series of synthetic intervals is proposed and visualized.

1) *Experiment on Aliasing Propensity*: In Fig. 3, four different pairs of intervals $\{\bar{a}, \bar{b}\}$ are considered where all pairs have an intersection of equal size. The similarity results for the pairs using the three SMs are shown in Table III. The S_J and S_D measures are subject to *aliasing*, providing the same similarity of 0.15 and 0.26, respectively, for all pairs.

Indeed, both measures provide—unexpectedly—identical similarities for pairs of intervals when the size of the union of

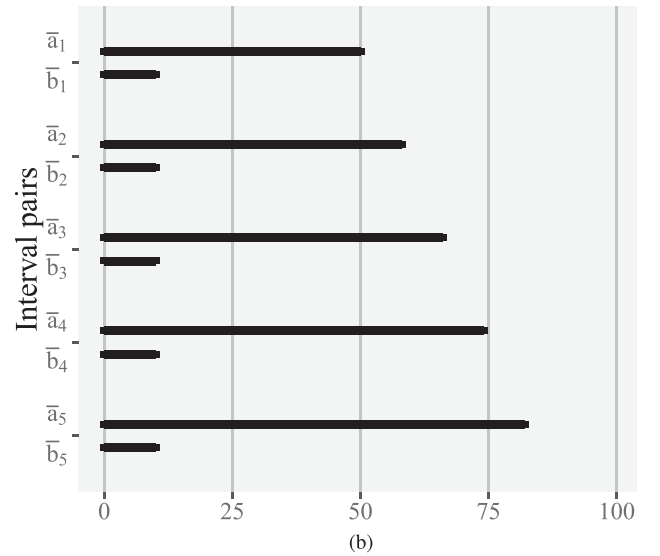
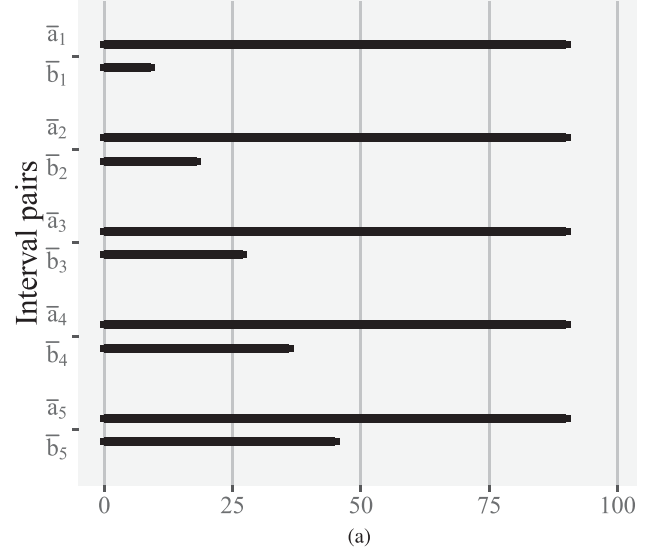


Fig. 4. Interval pairs where one interval is a complete subset of the other. (a) Increasing interval overlapping. (b) Increasing interval nonoverlapping.

their nonoverlapping segments remains constant. On the contrary, the S_{S_h} measure with both minimum and product t -norms ($S_{S_h \wedge}$ and $S_{S_h \Pi}$) yields a different degree of similarity for all cases, thereby exhibiting its *aliasing*-free and robust behavior. The reason is that the S_{S_h} measure captures the changes in the size of both input intervals as compared to S_J and S_D measures, which eventually affects their reciprocal subsethood and the overall similarity. Note that as shown in Theorem 11, the results of the $S_{S_h \wedge}$ measure are bounded by the S_J and S_D measures.

2) *Experiment With Interval Pairs When One Interval is a Complete Subset of the Other*: Two separate cases are considered with different sets of interval pairs, where in both cases, one interval is a complete subset of the other: an increasing degree of overlap between both intervals by increasing the size of the smaller interval, and decreasing the degree of overlap by increasing the size of the larger interval.

Fig. 4(a) shows five interval pairs with $\bar{b} \subset \bar{a}$, where \bar{b} covers 10%, 20%, 30%, 40%, and 50% of \bar{a} . On the contrary, in

TABLE IV
SIMILARITY RESULTS FOR THE INTERVAL PAIRS AS SHOWN IN FIG. 4

Interval pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	Subset by
$\{\bar{a}_1, \bar{b}_1\}$	0.10	0.18	0.10	0.10	10%
$\{\bar{a}_2, \bar{b}_2\}$	0.20	0.33	0.20	0.20	20%
$\{\bar{a}_3, \bar{b}_3\}$	0.30	0.46	0.30	0.30	30%
$\{\bar{a}_4, \bar{b}_4\}$	0.40	0.57	0.40	0.40	40%
$\{\bar{a}_5, \bar{b}_5\}$	0.50	0.67	0.50	0.50	50%

(a) Increasing interval overlapping

Interval pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	Subset by
$\{\bar{a}_1, \bar{b}_1\}$	0.20	0.33	0.20	0.20	20%
$\{\bar{a}_2, \bar{b}_2\}$	0.172	0.294	0.172	0.172	17.2%
$\{\bar{a}_3, \bar{b}_3\}$	0.152	0.263	0.152	0.152	15.2%
$\{\bar{a}_4, \bar{b}_4\}$	0.135	0.238	0.135	0.135	13.5%
$\{\bar{a}_5, \bar{b}_5\}$	0.122	0.217	0.122	0.122	12.2%

(b) Increasing interval nonoverlapping

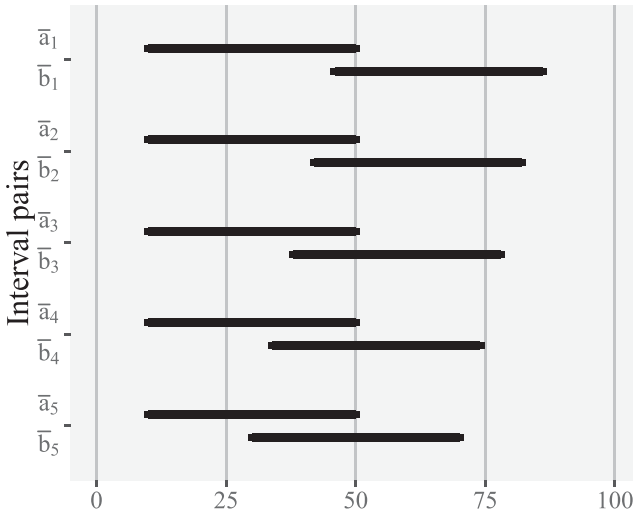


Fig. 5. Interval pairs with equal size and equal ratio of overlap.

Fig. 4(b), another five interval pairs are presented where the degree of overlap (in percent) is gradually decreased by increasing the size of \bar{a} . Table IV(a) and (b) shows the similarity for all pairs with all three SMs. Note that for all pairs, the subsethood of \bar{b} is 1, whereas the subsethood of \bar{a} depends on the size of \bar{a} and \bar{b} , i.e., $\frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|}$. Intuitively, their mutual similarity should be at most $\frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|}$ for each pair. From the results, we see that both S_{S_h} (with minimum and product t -norms) and S_J measures perform according to the intuition, whereas the S_D measure exceeds this expected limit.

3) *Experiment With Interval Pairs of Equal Size and Equal Overlapping Ratio:* In Fig. 5, five interval pairs are shown where the intervals are of equal size and the size of their intersection is varied to 10%, 20%, 30%, 40%, and 50% of their size. Table V provides the results for all pairs using the three SMs. In all pairs, the subsethood is equal, and it is intuitive to expect the similarity to be the same as this subsethood (as the intervals are both of equal size). In this case, the S_{S_h} with minimum t -norm ($S_{S_h \wedge}$) and S_D measures follow the intuition, whereas the S_{S_h} with

TABLE V
SIMILARITY RESULTS FOR THE INTERVAL PAIRS AS SHOWN IN FIG. 5

Interval pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	Intersecting by
$\{\bar{a}_1, \bar{b}_1\}$	0.05	0.10	0.10	0.01	10%
$\{\bar{a}_2, \bar{b}_2\}$	0.11	0.20	0.20	0.04	20%
$\{\bar{a}_3, \bar{b}_3\}$	0.18	0.30	0.30	0.09	30%
$\{\bar{a}_4, \bar{b}_4\}$	0.25	0.40	0.40	0.16	40%
$\{\bar{a}_5, \bar{b}_5\}$	0.33	0.50	0.50	0.25	50%

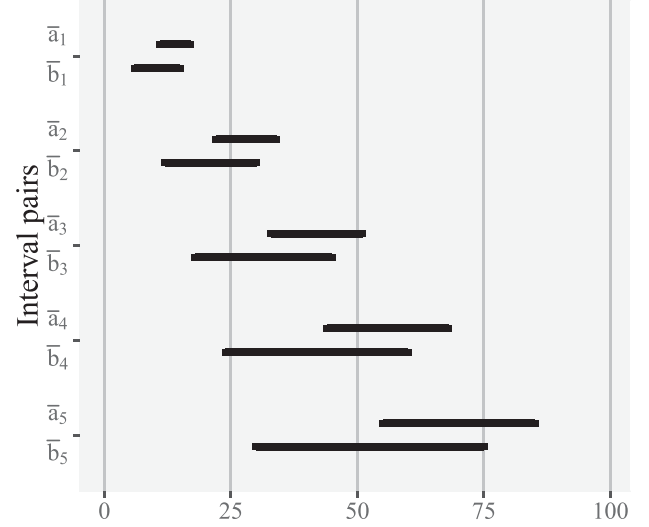


Fig. 6. Interval pairs used to demonstrate the invariance of SMs.

TABLE VI
ALL THREE SMs DISPLAY INVARIANT SIMILARITY
FOR THE INTERVAL PAIRS IN FIG. 6

Interval pair	S_J	S_D	$S_{S_h \wedge}$	$S_{S_h \Pi}$	Multiplied by n
$\{\bar{a}_1, \bar{b}_1\}$	0.36	0.53	0.44	0.29	1
$\{\bar{a}_2, \bar{b}_2\}$	0.36	0.53	0.44	0.29	2
$\{\bar{a}_3, \bar{b}_3\}$	0.36	0.53	0.44	0.29	3
$\{\bar{a}_4, \bar{b}_4\}$	0.36	0.53	0.44	0.29	4
$\{\bar{a}_5, \bar{b}_5\}$	0.36	0.53	0.44	0.29	5

product t -norm ($S_{S_h \Pi}$) and S_J measures yield a lower degree of similarity.

4) *Experiment on Scaling Invariance:* Five pairs of intervals are shown in Fig. 6, where both endpoints of \bar{a} and \bar{b} are gradually multiplied by a factor $n \in \{2, 3, 4, 5\}$ to generate new interval pairs. The degree of subsethood stays the same across the pairs. Adapting the definition from [53], an SM is *invariant* if its similarity output remains constant regardless of scaling the interval endpoints by a factor. Table VI shows the similarity for all pairs using the three SMs, where n is the factor applied to the interval endpoints. The results demonstrate the *scaling invariance* property for the given pairs of intervals for all SMs.

5) *Experiment on Increased Overlap Linearly:* Adapting the definition from [53], an SM on intervals is *linear* if its similarity output varies linearly as to a linear change in the size of the overlap/intersection of the intervals. In Fig. 7(a), the overlap between two intervals of equal size is gradually increased in 10% steps. The corresponding similarity outputs for the pairs

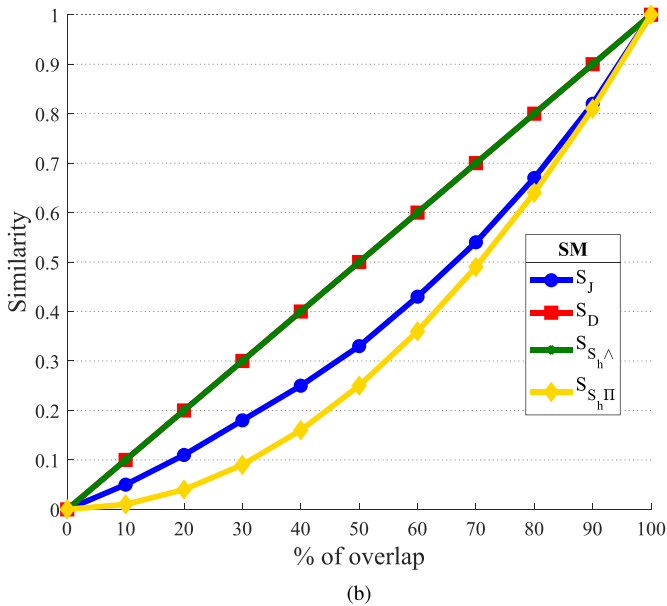
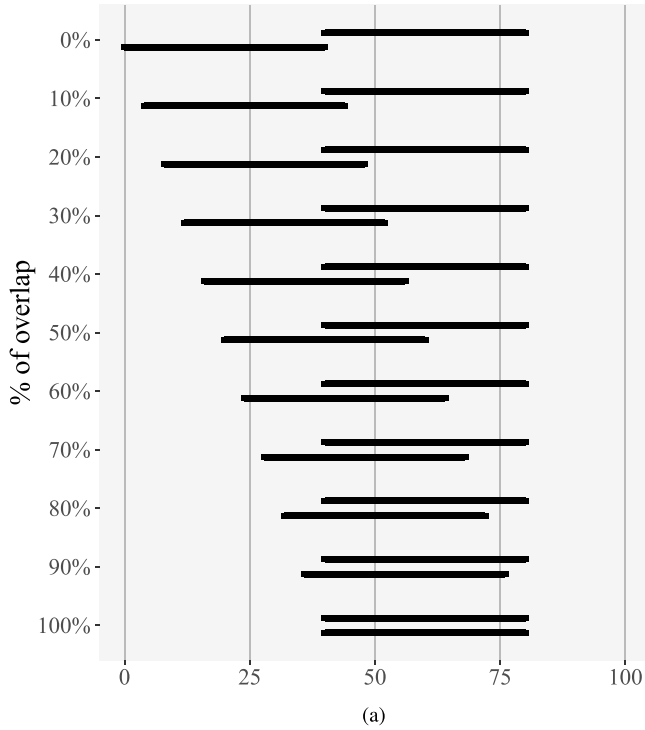


Fig. 7. Displaying linearity by the S_D and $S_{S_h \wedge}$ measures, and convexity by the S_J and $S_{S_h \Pi}$ measures, as shown in (b), for rising overlap when intervals are of equal size, as shown in (a).

and all three SMs are shown graphically in Fig. 7(b). Results show that all three SMs provide higher similarity for increased overlap. Particularly, S_{S_h} with minimum t -norm ($S_{S_h \wedge}$) and S_D measures exhibit *linearity* in the similarity results, whereas S_{S_h} with product t -norm ($S_{S_h \Pi}$) and S_J measures display *convexity* (similarity increases rapidly with an increase in the size of the intersection).

In summary, case 1 shows that S_J and S_D measures are subject to *aliasing* and return the same similarity for very

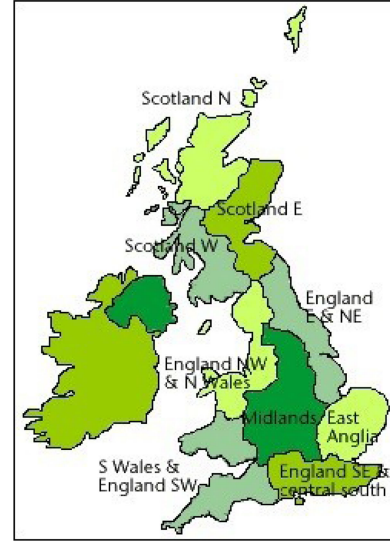


Fig. 8. Standard areas (districts) of the UK [54].

different interval pairs whenever the union of their nonoverlapping remains constant (regardless of any changes in their size). In case 2, when one interval is a complete subset of the other, proportionately increasing or decreasing their degree of overlap results in the S_{S_h} and S_J measures returning sensible results, whereas the S_D measure overestimates the similarity. In case 3, S_{S_h} with minimum t -norm ($S_{S_h \wedge}$) and S_D measures meet the expectation when intervals are of equal size and their intersection is changed proportionately, whereas the S_J measure underestimates the similarity. In case 4, we multiply interval end-points by a factor, maintaining the same degree of sub-sethood. Here, all three measures show *scaling invariance* in the results as expected. Finally, in case 5, for linear increases in the size of overlap between the intervals of equal size, S_{S_h} with minimum t -norm ($S_{S_h \wedge}$) and S_D measures exhibit *linearity* in the similarity results, whereas the S_{S_h} with product t -norm ($S_{S_h \Pi}$) and S_J measures display *convexity*. These experiments demonstrate that the proposed SM shows behavior in line with expectation, which is not the case for the other, commonly used measures.

B. Real-World Example

In this part, we have used a real-world dataset to review the behavior of the three SMs [S_{S_h} with minimum (\wedge) and product (Π) t -norms, S_J , and S_D], particularly in respect to the *aliasing* issue. The dataset used for this demonstration is the temperature data of different standard areas (districts) of the UK for the year 2016 from the UK Met Office. The entire dataset is available in [54]. The map of the standard areas used by the UK Met Office is shown in Fig. 8. We have used temperature data of nine standard areas: Scotland North (SN), Scotland East (SE), Scotland West (SW), England East and North East (EEN), England Northwest and Wales North (ENW), Midlands (ML), East Anglia (EA), England South–West and Wales South (ESW), and England South–East and Central South (ESC) for four different seasons in 2016.

TABLE VII
SEASONAL TEMPERATURE IN STANDARD AREAS OF THE UK IN 2016

Area	Acronym	Seasons							
		Winter		Spring		Summer		Autumn	
		Min	Max	Min	Max	Min	Max	Min	Max
Scotland N	SN	0.5	6.0	2.7	9.8	9.1	15.8	4.8	11.2
Scotland E	SE	0.4	6.2	2.4	10.3	9.1	16.9	4.4	11.5
Scotland W	SW	1.9	7.2	3.2	11.1	10.0	17.3	5.6	12.0
England E and NE	EEN	2.4	8.5	3.6	11.6	10.9	19.5	6.6	13.5
England NW and Wales N	ENW	3.1	8.6	3.9	11.8	11.4	18.7	6.5	13.1
Midlands	ML	3.3	9.4	3.9	12.6	11.7	20.6	6.9	14.1
East Anglia	EA	3.8	10.0	4.5	13.2	12.3	22.1	7.7	15.5
England SW and Wales S	ESW	4.5	9.8	4.4	12.7	11.9	19.5	7.3	14.2
England SE and Central S	ESC	4.3	10.2	4.7	13.5	12.4	21.6	7.5	15.4

TABLE VIII
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE S_J MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9014	0.7778	0.5287	0.5663	0.4624	0.3271	0.4149	0.3491	1	0.8987	0.7857	0.6966	0.6484	0.596	0.5048	0.54	0.4722
SE	0.9014	1	0.7763	0.5385	0.5747	0.4742	0.3423	0.4286	0.3636	0.8987	1	0.8161	0.7283	0.6809	0.6275	0.537	0.5728	0.5045
SW	0.7778	0.7763	1	0.6835	0.7333	0.6	0.4343	0.5465	0.4592	0.7857	0.8161	1	0.8929	0.8372	0.766	0.66	0.7053	0.6214
EEN	0.5287	0.5385	0.6835	1	0.9286	0.88	0.6517	0.8158	0.6818	0.6966	0.7283	0.8929	1	0.939	0.8556	0.7396	0.7912	0.697
ENW	0.5663	0.5747	0.7333	0.9286	1	0.8158	0.6	0.7532	0.6292	0.6484	0.6809	0.8372	0.939	1	0.908	0.7849	0.8409	0.7396
ML	0.4624	0.4742	0.6	0.88	0.8158	1	0.7442	0.9315	0.7765	0.596	0.6275	0.766	0.8556	0.908	1	0.871	0.9318	0.8229
EA	0.3271	0.3423	0.4343	0.6517	0.6	0.7442	1	0.7927	0.9625	0.5048	0.537	0.66	0.7396	0.7849	0.871	1	0.9318	0.9444
ESW	0.4149	0.4286	0.5465	0.8158	0.7532	0.9315	0.7927	1	0.8272	0.54	0.5728	0.7053	0.7912	0.8409	0.9318	0.9318	1	0.8791
ESC	0.3491	0.3636	0.4592	0.6818	0.6292	0.7765	0.9625	0.8272	1	0.4722	0.5045	0.6214	0.697	0.7396	0.8229	0.9444	0.8791	1

TABLE IX
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE S_D MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9481	0.875	0.6917	0.7231	0.6324	0.493	0.5865	0.5175	1	0.9467	0.88	0.8212	0.7867	0.7468	0.6709	0.7013	0.6415
SE	0.9481	1	0.8741	0.7	0.7299	0.6434	0.5101	0.6	0.5333	0.9467	1	0.8987	0.8428	0.8101	0.7711	0.6988	0.7284	0.6707
SW	0.875	0.8741	1	0.812	0.8462	0.75	0.6056	0.7068	0.6294	0.88	0.8987	1	0.9434	0.9114	0.8675	0.7952	0.8272	0.7665
EEN	0.6917	0.7	0.812	1	0.963	0.9362	0.7891	0.8986	0.8108	0.8212	0.8428	0.9434	1	0.9686	0.9222	0.8503	0.8834	0.8214
ENW	0.7231	0.7299	0.8462	0.963	1	0.8986	0.75	0.8593	0.7724	0.7867	0.8101	0.9114	0.9686	1	0.9518	0.8795	0.9136	0.8503
ML	0.6324	0.6434	0.75	0.9362	0.8986	1	0.8533	0.9645	0.8742	0.7468	0.7711	0.8675	0.9222	0.9518	1	0.931	0.9647	0.9029
EA	0.493	0.5101	0.6056	0.7891	0.75	0.8533	1	0.8844	0.9809	0.6709	0.6988	0.7952	0.8503	0.8795	0.931	1	0.9647	0.9714
ESW	0.5865	0.6	0.7068	0.8986	0.8593	0.9645	0.8844	1	0.9054	0.7013	0.7284	0.8272	0.8834	0.9136	0.9647	0.9647	1	0.9357
ESC	0.5175	0.5333	0.6294	0.8108	0.7724	0.8742	0.9809	0.9054	1	0.6415	0.6707	0.7665	0.8214	0.8503	0.9029	0.9714	0.9357	1

TABLE X
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE $S_{S_h \wedge}$ MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9014	0.875	0.6667	0.7121	0.5972	0.4487	0.5652	0.4684	1	0.8987	0.8354	0.775	0.7468	0.6782	0.6092	0.6506	0.5795
SE	0.9014	1	0.831	0.6901	0.7042	0.6389	0.4872	0.5915	0.5063	0.8987	1	0.8987	0.8375	0.8101	0.7356	0.6667	0.7108	0.6364
SW	0.875	0.831	1	0.7826	0.8333	0.7083	0.5513	0.6812	0.5696	0.8354	0.8987	1	0.9375	0.9114	0.8276	0.7586	0.8072	0.7273
EEN	0.6667	0.6901	0.7826	1	0.942	0.9167	0.7436	0.8986	0.7595	0.775	0.8375	0.9375	1	0.9625	0.8851	0.8161	0.8675	0.7841
ENW	0.7121	0.7042	0.8333	0.942	1	0.8611	0.6923	0.8406	0.7089	0.7468	0.8101	0.9114	0.9625	1	0.908	0.8391	0.8916	0.8068
ML	0.5972	0.6389	0.7083	0.9167	0.8611	1	0.8205	0.9444	0.8354	0.6782	0.7356	0.8276	0.8851	0.908	1	0.931	0.9425	0.8977
EA	0.4487	0.4872	0.5513	0.7436	0.6923	0.8205	1	0.8333	0.9747	0.6092	0.6667	0.7586	0.8161	0.8391	0.931	1	0.9425	0.9659
ESW	0.5652	0.5915	0.6812	0.8986	0.8406	0.9444	0.8333	1	0.8481	0.6506	0.7108	0.8072	0.8675	0.8916	0.9425	0.9425	1	0.9091
ESC	0.4684	0.5063	0.5696	0.7595	0.7089	0.8354	0.9747	0.8481	1	0.5795	0.6364	0.7273	0.7841	0.8086	0.8977	0.9659	0.9091	1

The intervals are constructed by taking the minimum and the maximum temperature of each area for every season, as shown in Table VII. We have applied the S_{S_h} [with minimum (\wedge) and product (Π) t -norms], S_J and S_D measures to estimate similarities between all possible pairs of the nine areas for all seasons; here, the results for autumn and spring seasons are reported, which are shown in Tables VIII–XI. We do not include results for other two seasons—winter and summer—as they follow the same pattern.

Regarding similarity based on seasonal temperature, one can expect higher similarity for pairs of areas geographically close to each other (i.e., located in the same latitude), whereas expecting lower similarity for remote areas as it is generally recognized that latitude (i.e., distance from the equator) is an important factor affecting variations in temperature. Hence, in this real-world example, the districts in the Northern UK (e.g., SE and SW) are likely to be more similar as they are in the South (e.g., ML and EA). By the same token, a district in the north will be highly

TABLE XI
SIMILARITY RESULTS FOR TEMPERATURE DATA USING THE $S_{S_h \Pi}$ MEASURE

Area	Autumn									Spring								
	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC	SN	SE	SW	EEN	ENW	ML	EA	ESW	ESC
SN	1	0.9014	0.7656	0.4792	0.523	0.4013	0.2454	0.3444	0.2708	1	0.8987	0.7766	0.6768	0.6206	0.5635	0.4548	0.4948	0.4163
SE	0.9014	1	0.7661	0.4901	0.5335	0.4139	0.2607	0.3601	0.2853	0.8987	1	0.8077	0.7103	0.6563	0.596	0.4895	0.5309	0.4511
SW	0.7656	0.7661	1	0.6603	0.7161	0.5645	0.3704	0.5002	0.4005	0.7766	0.8077	1	0.89	0.8306	0.7543	0.6338	0.6846	0.5892
EEN	0.4792	0.4901	0.6603	1	0.9278	0.8768	0.625	0.8074	0.6604	0.6768	0.7103	0.89	1	0.9381	0.8519	0.7243	0.7807	0.6763
ENW	0.523	0.5335	0.7161	0.9278	1	0.8089	0.5664	0.7387	0.6015	0.6206	0.6563	0.8306	0.9381	1	0.908	0.7754	0.8351	0.7251
ML	0.4013	0.4139	0.5645	0.8768	0.8089	1	0.7293	0.9308	0.7658	0.5635	0.596	0.7543	0.8519	0.908	1	0.8668	0.9312	0.8152
EA	0.2454	0.2607	0.3704	0.625	0.5664	0.7293	1	0.785	0.9622	0.4548	0.4895	0.6338	0.7243	0.7754	0.8668	1	0.9312	0.9437
ESW	0.3444	0.3601	0.5002	0.8074	0.7387	0.9308	0.785	1	0.8235	0.4948	0.5309	0.6846	0.7807	0.8351	0.9312	0.9312	1	0.8762
ESC	0.2708	0.2853	0.4005	0.6604	0.6015	0.7658	0.9622	0.8235	1	0.4163	0.4511	0.5892	0.6763	0.7251	0.8152	0.9437	0.8762	1

dissimilar to a district in the south. Generally, all three SMs appear to meet up with such expectations. For instance, considering autumn temperatures between SN and the other eight areas from geographically close to far off, three SMs appear to generate higher to lower similarity as temperature-range differs more.

However, in several cases, the S_J and S_D measures suffer from the *aliasing*—returning the same similarity for different ranges of temperature. As mentioned earlier, both measures provide identical similarity for pairs of areas as long as the size of the intersection and union of their temperature-range remain constant. Tables VIII and IX show that they produce the same similarity (i.e., 0.7396 and 0.8503, respectively) for pairs of areas, (EEN versus EA) and (ENW versus ESC) in spring, even though they have separate temperature ranges. Analogously, we also notice identical similarities for the pairs, (SW versus ML) and (EA versus ENW), and the pairs (ML versus ENW) and (ESW versus EEN) in autumn. In contrast, the S_{S_h} measure with minimum and product t -norms (i.e., $S_{S_h \wedge}$ and $S_{S_h \Pi}$), responding to variation in temperature range, yields a distinct outcome for all of these pairs, as shown in Tables X and XI.

This real-world example showing the *aliasing* inherent to the S_J and S_D measures highlights the potential for misleading inference in practice. In particular, while clustering the areas with respect to temperature, some areas may be placed in the same group albeit having different ranges of temperature. This can be an issue when we compare real-world interval data for the purpose of grouping, ranking, and decision making. Therefore, an SM that avoids this *aliasing*—such as the proposed measure, is desirable.

V. CONCLUSION

The contributions of this article centered on to identification and articulation of potential shortcomings—specifically *aliasing*—affecting popular set-theoretic SMs. The latter have the potential to give misleading results, in turn affecting real-world applications dependent on the robust comparison of intervals. In order to address this limitation, this article developed the underpinnings of robust SMs by putting forward five new axioms that complement the traditional axioms associated with the most common SMs. Building on this set of nine axioms, this article established a new set-theoretic SM for intervals based on their bidirectional subthood. The new SM was shown to avoid shortcomings, such as *aliasing*, while delivering intuitive results (e.g., being bounded above and below by Jaccard and Dice SMs),

facilitating its potential use in real-world applications. As part of the development of the new SM, the article also put forward the definition of subthood for intervals and proofs of its mathematical properties (see Appendix). Finally, a corresponding dissimilarity or DM was derived from the new SM, which was also proven to be a metric.

At an experimental level, the article provided a detailed investigation contrasting the behavior of the proposed SM vis-a-vis the Jaccard and Dice SMs—using both synthetic and real-world interval-valued data. The exhaustive analyses confirmed that the new measure exhibits desirable behavior while maintaining all essential features of an SM. In particular, the new SM is resilient to *aliasing* and provides desirable results in respect to all key features (e.g., *linearity* and *scaling invariance*), whereas popular SMs are shown to produce counter-intuitive results in some cases.

In the future, we plan to use this new measure for assessing similarity between discontinuous intervals and to develop corresponding extensions of the SM for type-1, interval-valued, and type-2 fuzzy sets using the α -cut decomposition representation. We also aim to explore it for capturing the mutual agreement of interval-valued evidence for aggregation as well as clustering and classification of real-world interval-valued datasets.

APPENDIX

DEFINITION OF SUBTHEOOD FOR CLOSED INTERVALS

Building on the axiomatic definitions of subthood in the literature (see Section II-C), the key properties of subthood for intervals are captured in the following theorem.

Theorem 14: A real-valued function $S_h : \bar{a} \times \bar{b} \rightarrow [0, 1]$ captures subthood for intervals, if it follows the following properties.

- 1) $S_h(\bar{a}, \bar{b}) = 1$ if and only if $\bar{a} \subseteq \bar{b}$.
- 2) $S_h(\bar{a}, \bar{b}) = 0$ if and only if \bar{a} and \bar{b} are disjoint.
- 3) if $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$, then $S_h(\bar{c}, \bar{a}) \leq S_h(\bar{b}, \bar{a})$ and $S_h(\bar{c}, \bar{a}) \leq S_h(\bar{c}, \bar{b})$.

Proof: Proofs for all the properties (1)–(3) are as follows.

- 1) If $\bar{a} \subseteq \bar{b}$, then $|\bar{a} \cap \bar{b}| = |\bar{a}|$. Thus, $S_h(\bar{a}, \bar{b}) = \frac{|\bar{a} \cap \bar{b}|}{|\bar{a}|} = \frac{|\bar{a}|}{|\bar{a}|} = 1$. Again, $S_h(\bar{a}, \bar{b}) = 1$ only when \bar{a} is included in \bar{b} .
- 2) If \bar{a} and \bar{b} are disjoint, then $|\bar{a} \cap \bar{b}| = 0$. Thus, $S_h(\bar{a}, \bar{b}) = 0$. Contrarily, $S_h(\bar{a}, \bar{b}) = 0$ when \bar{a} and \bar{b} are not overlapping.
- 3) If $\bar{a} \subseteq \bar{b} \subseteq \bar{c}$, then $|\bar{a}| \leq |\bar{b}| \leq |\bar{c}|$. Thus, $S_h(\bar{c}, \bar{a}) = \frac{|\bar{c} \cap \bar{a}|}{|\bar{c}|} = \frac{|\bar{a}|}{|\bar{c}|}$ and $S_h(\bar{b}, \bar{a}) = \frac{|\bar{b} \cap \bar{a}|}{|\bar{b}|} = \frac{|\bar{a}|}{|\bar{b}|}$. Since $|\bar{b}| \leq |\bar{c}|$, it follows

that $\frac{1}{|c|} \leq \frac{1}{|b|}$. Hence, $\frac{|a|}{|c|} \leq \frac{|a|}{|b|}$, implying $S_h(\bar{c}, \bar{a}) \leq S_h(\bar{b}, \bar{a})$. Moreover, $S_h(\bar{c}, \bar{b}) = \frac{|c \cap b|}{|c|} = \frac{|b|}{|c|}$. Since $|a| \leq |b|$, it follows that $\frac{|a|}{|c|} \leq \frac{|b|}{|c|}$, proving that $S_h(\bar{c}, \bar{a}) \leq S_h(\bar{c}, \bar{b})$. ■

REFERENCES

- [1] A. Tversky, "Features of similarity," *Psychol. Rev.*, vol. 84, no. 4, pp. 327–352, 1977.
- [2] I. Beg and S. Ashraf, "Similarity measures for fuzzy sets," *Appl. Comput. Math.*, vol. 8, no. 2, pp. 192–202, 2009.
- [3] M. Galar, J. Fernández, G. Beliakov, and H. Bustince, "Interval-valued fuzzy sets applied to stereo matching of color images," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1949–1961, Jul. 2011.
- [4] C. Wagner, S. Miller, J. M. Garibaldi, D. T. Anderson, and T. C. Havens, "From interval-valued data to general type-2 fuzzy sets," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 2, pp. 248–269, Apr. 2015.
- [5] D. S. Guru, B. B. Kiranagi, and P. Nagabhushan, "Multivalued type proximity measure and concept of mutual similarity value useful for clustering symbolic patterns," *Pattern Recognit. Lett.*, vol. 25, no. 10, pp. 1203–1213, 2004.
- [6] Y. Ren, Y.-H. Liu, J. Rong, and R. Dew, "Clustering interval-valued data using an overlapped interval divergence," in *Proc. 8th Australas. Data Mining Conf.*, 2009, pp. 35–42.
- [7] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bull. de la Soci. Vaudoise Des Sci. Naturelles*, vol. 44, pp. 223–270, 1908.
- [8] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [9] S. Kabir, C. Wagner, T. C. Havens, D. T. Anderson, and U. Aickelin, "Novel similarity measure for interval-valued data based on overlapping ratio," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2017, pp. 1–6.
- [10] D. Szymkiewicz, "Une contribution statistique à la géographie floristique," *Acta Societatis Botanicorum Poloniae*, vol. 11, no. 3, pp. 249–265, 1934.
- [11] Y. Otsuka, "The faunal character of the Japanese pleistocene marine mollusca, as evidence of climate having become colder during the pleistocene in Japan," *Bull. Biogeographical Soc. Jpn.*, vol. 6, no. 16, pp. 165–170, 1936.
- [12] R. R. Sokal and P. H. A. Sneath, "Principles of numerical taxonomy," *Princ. Numer. Taxonomy*, vol. 13, pp. 106–108, 1963.
- [13] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationship," *Univ. Kansas Sci. Bull.*, vol. 28, pp. 1409–1438, 1958.
- [14] J. McCulloch, C. Wagner, and U. Aickelin, "Analysing fuzzy sets through combining measures of similarity and distance," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2014, pp. 155–162.
- [15] K. H. Rosen, *Discrete Mathematics and Its Applications*, 7th ed. New York, NY, USA: McGraw-Hill, 2012.
- [16] T. C. Havens, D. T. Anderson, C. Wagner, H. Deilamsalehy, and D. Wonnacott, "Fuzzy integrals of crowd-sourced intervals using a measure of generalized accord," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2013, pp. 1–8.
- [17] T. C. Havens, D. T. Anderson, and C. Wagner, "Data-informed fuzzy measures for fuzzy integration of intervals and fuzzy numbers," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 5, pp. 1861–1875, Oct. 2015.
- [18] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York, NY, USA: Academic, 1980.
- [19] C. Wagner, D. T. Anderson, and T. C. Havens, "Generalization of the fuzzy integral for discontinuous interval-valued and non-convex interval fuzzy set-valued inputs," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2013, pp. 1–8.
- [20] G. Beliakov and S. James, "A penalty-based aggregation operator for non-convex intervals," *Knowl.-Based Syst.*, vol. 70, pp. 335–344, 2014.
- [21] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [22] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ, USA: Prentice-Hall, vol. 8, pp. 338–353, 2001.
- [23] C. Wagner, S. Miller, and J. M. Garibaldi, "Similarity based applications for data-driven concept and word models based on type-1 and type-2 fuzzy sets," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2013, pp. 1–9.
- [24] H. T. Nguyen and K. Vladik, "Computing degrees of subsethood and similarity for interval-valued fuzzy sets: Fast algorithms," in *Proc. 9th Int. Conf. Intell. Technol.*, 2008, pp. 47–55.
- [25] A. Stachowiak, P. Żywica, K. Dyczkowski, and A. Wójtowicz, "An interval-valued fuzzy classifier based on an uncertainty-aware similarity measure," in *Intelligent Systems' 2014*. Berlin, Germany: Springer, 2014, pp. 741–751.
- [26] D. Wu and J. M. Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets," *Inf. Sci.*, vol. 179, no. 8, pp. 1169–1192, 2009.
- [27] J. McCulloch, C. Wagner, and U. Aickelin, "Extending similarity measures of interval type-2 fuzzy sets to general type-2 fuzzy sets," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2013, pp. 1–8.
- [28] J. Ye, "Multicriteria decision-making method using the dice similarity measure between expected intervals of trapezoidal fuzzy numbers," *J. Decis. Syst.*, vol. 21, no. 4, pp. 307–317, 2012.
- [29] J. Ye, "The dice similarity measure between generalized trapezoidal fuzzy numbers based on the expected interval and its multicriteria group decision-making method," *J. Chin. Inst. Ind. Eng.*, vol. 29, no. 6, pp. 375–382, 2012.
- [30] L. Baccour and R. I. John, "Experimental analysis of crisp similarity and distance measures," in *Proc. 6th Int. Conf. Soft Comput. Pattern Recognit.*, 2014, pp. 96–100.
- [31] A. Gardner, J. Kanno, C. A. Duncan, and R. Selmic, "Measuring distance between unordered sets of different sizes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 137–143.
- [32] Y. Li, K. Qin, and X. He, "Relations among similarity measure, subsethood measure and fuzzy entropy," *Int. J. Comput. Intell. Syst.*, vol. 6, no. 3, pp. 411–422, 2013.
- [33] B. Kosko, "Fuzzy entropy and conditioning," *Inf. Sci.*, vol. 40, no. 2, pp. 165–174, 1986.
- [34] W. Bandler and L. Kohout, "Fuzzy power sets and fuzzy implication operators," *Fuzzy Sets Syst.*, vol. 4, no. 1, pp. 13–30, 1980.
- [35] L. Kitainik, "Fuzzy inclusions and fuzzy dichotomous decision procedures," in *Optimization Models Using Fuzzy Sets and Possibility Theory*. Berlin, Germany: Springer, 1987, pp. 154–170.
- [36] D. Sinha and E. R. Dougherty, "Fuzzification of set inclusion: Theory and applications," *Fuzzy Sets Syst.*, vol. 55, no. 1, pp. 15–42, 1993.
- [37] V. R. Young, "Fuzzy subsethood," *Fuzzy Sets Syst.*, vol. 77, no. 3, pp. 371–384, 1996.
- [38] J. Fan, W. Xie, and J. Pei, "Subsethood measure: New definitions," *Fuzzy Sets Syst.*, vol. 106, no. 2, pp. 201–209, 1999.
- [39] H. Bustince, V. Mohedano, E. Barrenechea, and M. Pagola, "Definition and construction of fuzzy DI-subsethood measures," *Inf. Sci.*, vol. 176, no. 21, pp. 3190–3231, 2006.
- [40] H. Bustince, "Indicator of inclusion grade for interval-valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets," *Int. J. Approx. Reasoning*, vol. 23, no. 3, pp. 137–209, 2000.
- [41] W. Zeng and P. Guo, "Normalized distance, similarity measure, inclusion measure and entropy of interval-valued fuzzy sets and their relationship," *Inf. Sci.*, vol. 178, no. 5, pp. 1334–1342, 2008.
- [42] I. K. Vlachos and G. D. Sergiadis, "Subsethood, entropy, and cardinality for interval-valued fuzzy sets: an algebraic derivation," *Fuzzy Sets Syst.*, vol. 158, no. 12, pp. 1384–1396, 2007.
- [43] Z. Takáč, M. Minárová, J. Montero, E. Barrenechea, J. Fernandez, and H. Bustince, "Interval-valued fuzzy strong S-subsethood measures, interval-entropy and P-interval-entropy," *Inf. Sci.*, vol. 432, pp. 97–115, 2018.
- [44] P. Grzegorzewski, "On possible and necessary inclusion of intuitionistic fuzzy sets," *Inf. Sci.*, vol. 181, no. 2, pp. 342–350, 2011.
- [45] Z. Takáč, "Inclusion and subsethood measure for interval-valued fuzzy sets and for continuous type-2 fuzzy sets," *Fuzzy Sets Syst.*, vol. 244, no. 1, pp. 106–120, 2013.
- [46] W. Zeng and H. Li, "Inclusion measures, similarity measures, and the fuzziness of fuzzy sets and their relations," *Int. J. Intell. Syst.*, vol. 21, no. 6, pp. 639–653, 2006.
- [47] L. Xuecheng, "Entropy, distance measure and similarity measure of fuzzy sets and their relations," *Fuzzy Sets Syst.*, vol. 52, no. 3, pp. 305–318, 1992.
- [48] C. P. Pappis and N. I. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy values," *Fuzzy Sets Syst.*, vol. 56, no. 2, pp. 171–174, 1993.
- [49] H. Bustince, E. Barrenechea, and M. Pagola, "Image thresholding using restricted equivalence functions and maximizing the measures of similarity," *Fuzzy Sets Syst.*, vol. 158, no. 5, pp. 496–516, 2007.
- [50] H. Bustince, E. Barrenechea, and M. Pagola, "Relationship between restricted dissimilarity functions, restricted equivalence functions and normal E_N -functions: Image thresholding invariant," *Pattern Recognit. Lett.*, vol. 29, no. 4, pp. 525–536, 2008.

- [51] H. Bustince, E. Barrenechea, and M. Pagola, "Restricted equivalence functions," *Fuzzy Sets Syst.*, vol. 157, no. 17, pp. 2333–2346, 2006.
- [52] D. M. Gabbay and J. Woods, *Handbook of the History of Logic: The Many Valued and Nonmonotonic Turn in Logic*. Amsterdam, The Netherlands: Elsevier, 2007.
- [53] R. E. Tulloss, "Assessment of similarity indices for undesirable properties and a new tripartite similarity index based on cost functions," in *Mycology in Sustainable Development: Expanding Concepts, Vanishing Borders*. Boone, NC, USA: Parkway Publishers, 1997, pp. 122–143.
- [54] U.K. Met Office, "2016 weather summaries." Accessed: Dec. 16, 2017. [Online]. Available: <https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-and-regional-series>



Shaily Kabir (S'17) received the B.Sc. and M.S. degrees in computer science and engineering from the University of Dhaka, Dhaka, Bangladesh in 2003 and 2005 respectively, and the master's degree in computer science (research) from Concordia University, Montreal, QC, Canada in 2012. Shaily is currently working toward the Ph.D. degree in computer science from the University of Nottingham, Nottingham, U.K.

Since 2015, she has been a Commonwealth Scholar. Her main research interests include robust comparison of uncertain data received from multiple data sources and their fusion.

Mrs. Kabir was the recipient of the IEEE Computational Intelligence Society Graduate Research Grant in 2018.



Christian Wagner (SM'13) received the Ph.D. degree in computer science from the University of Essex, Colchester, U.K., in 2009.

He is currently a Professor of computer science with the University of Nottingham, Nottingham, U.K., and the Founding Director of the Lab for Uncertainty in Data and Decision Making. In 2017, he was recognized as a Recognising Inspirational Scientists and Engineers Connector by Engineering and Physical Sciences Research Council, Swindon, U.K. His work ranges from decision support in cybersecurity

and environmental management to personalization and control in manufacturing. He has authored/coauthored more than 100 peer-reviewed articles, focusing on modeling and handling of uncertain data arising from heterogeneous data sources (e.g., stakeholders), with a particular emphasis on designing interpretable artificial intelligence-based decision support systems.

Dr. Wagner is an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS journal, the Chair of the IEEE Computational Intelligence Society (CIS) Technical Committee on Fuzzy Systems, an elected member-at-large of the IEEE CIS Administrative Committee for 2018–2020. He was the Special Sessions Chair at Fuzz-IEEE 2019 in New Orleans, USA, and is the General Co-Chair of Fuzz-IEEE 2021 in Luxembourg.



Timothy Craig Havens (SM'10) received the Ph.D. degree in electrical and computer engineering from the University of Missouri, Columbia, MO, USA, in 2010.

He is currently the William and Gloria Jackson Associate Professor of computer systems and an Associate Dean for Research with the College of Computing, Michigan Technological University, Houghton, MI, USA. He is the Director of the Institute of Computing and Cybersystems (ICC) and ICC Center for Data Sciences, and cofounded the East Hancock Institute for Feline Companions.

Dr. Havens was a recipient of the Best Paper Award at the 2012 IEEE International Conference on Fuzzy Systems, the IEEE Franklin V. Taylor Award for Best Paper at the 2011 IEEE Conference on Systems, Man, and Cybernetics, and the Best Student Paper Award from the *Western Journal of Nursing Research* in 2009. He was the General Co-Chair of 2019 IEEE International Conference on Fuzzy Systems and is an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS.



Derek T. Anderson (SM'13) received the Ph.D. degree in electrical and computer engineering from the University of Missouri, Columbia, MO, USA, in 2010.

He is currently an Associate Professor of electrical engineering and computer science with the University of Missouri and the Director of the Mizzou Information and Data Fusion Laboratory. His research include information fusion in computational intelligence for signal/image processing, computer vision, and geospatial applications.

Dr. Anderson coauthored the 2013 Best Student Paper in Automatic Target Recognition at Society of Photo-Optical Instrumentation Engineers (SPIE). He was the recipient of the Best Overall Paper Award at the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), in 2012, and the 2008 FUZZ-IEEE Best Student Paper Award. He was also the Program Co-Chair of FUZZ-IEEE 2019, an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS, and an Area Editor for the *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*.