New Similarity Measures Between Generalized Trapezoidal Fuzzy Numbers Using the Jaccard Index

Chao-Ming Hwang
Department of Applied Mathematics,
Chinese Culture University, Taipei, Taiwan

Miin-Shen Yang*
Department of Applied Mathematics,
Chung Yuan Christian University, Chung-Li 32023, Taiwan
msyang@math.cycu.edu.tw

Received 1 May 2013
Revised 29 July 2014

Similarity measures between generalized trapezoidal fuzzy numbers (GTFNs) are employed to indicate the degrees of similarity between GTFNs. Although several similarity measures of GTFNs have been proposed in the literature, none has considered using the Jaccard index. In general, the Jaccard index is a statistic used for comparing the similarity and diversity of sample sets. This paper presents a new similarity measure between GTFNs, which involves the Jaccard index. The proposed similarity measure is found to have better properties. Several examples are employed to compare the proposed measure with some existing methods. An experiment is performed using 15 sets of GTFNs to compare the proposed similarity measure with existing ones. Numerical results show that the proposed measure is more reasonable than those existing methods.

Keywords: Generalized trapezoidal fuzzy number; fuzzy data; similarity measure; Jaccard index.

1. Introduction

Since Zadeh\(^1\) proposed the concept of fuzzy sets, fuzziness has been widely employed to handle a type of uncertainty that is different from probability (randomness). For treating fuzziness, fuzzy data and fuzzy presentations are commonly used. In all these fuzzy types of presentations, generalized trapezoidal fuzzy numbers (GTFNs) are most widely used in linguistics, decision making, knowledge representation, medical diagnosis, control systems, databases, clustering, and so forth.\(^2,3\)

A similarity measure serves to provide the degree of similarity between two objects. Kaufman and Rousseeuw\(^4\) presented some examples to illustrate traditional similarity measure applied in hierarchical clustering. Pappis and Karacapilidis\(^5\) proposed three similarity measures for fuzzy sets. More different similarity measures between fuzzy sets have been proposed in the literature.\(^6-9\) GTFNs are commonly employed to present fuzzy data.
data for analyzing fuzziness in a fuzzy environment. They have been applied in various areas, such as, in linguistics, control, database systems and clustering. In the literature, Chen first presented a similarity measure between GTFNs according to the geometric distance. Hsieh and Chen proposed a similarity measure between GTFNs using the graded mean integration representation distance. Lee presented a similarity measure between GTFNs that involves $L_p$ norm. Chen and Chen proposed a similarity measure between GTFNs using the center of gravity to define a similarity for GTFNs. Yang et al. proposed a similarity measure between GTFNs by defining an exponential metric. Wei and Chen proposed a similarity measure between GTFNs, which combines the concepts of geometric distance, perimeter and height of generalized fuzzy numbers to calculate the degree of similarity between GTFNs. Xu et al. have recently proposed another new similarity by combining the concept of geometric distance and center of gravity distance of GTFNs.

The Jaccard index (see Jaccard and Real and Vargas) is a statistic used for comparing the similarity and diversity of sample sets. Although many similarity measures between GTFNs had been proposed in the literature, none has considered using the Jaccard index. In this paper, a new similarity measure between GTFNs is developed using the Jaccard index. Some examples are employed to compare the proposed similarity measure with several existing methods. Numerical results show that the proposed similarity measure is better than existing ones developed by Chen, Hsieh and Chen, Lee, Chen and Chen, Wei and Chen and Xu et al. It has been found that these existing measures cannot accurately calculate the degrees of similarity between two GTFNs in some situations. The proposed similarity measure can overcome such drawback.

The rest of this paper is organized as follows. Section 2 reviews briefly basic concepts of existing similarity measures proposed by Chen, Hsieh and Chen, Lee, Chen and Chen, Wei and Chen and Xu et al. In Sec. 3, the Jaccard index is defined and the new similarity measure between GTFNs developed using the Jaccard index is proposed. Some properties are also presented. In Sec. 4, some examples are illustrated and comparisons are made with the existing similarity measures. Finally, conclusions are stated in Sec. 5.

2. Similarity Measures Between Generalized Trapezoidal Fuzzy Numbers

A fuzzy number $\tilde{A}$ is defined as a convex fuzzy set of real numbers with its membership function $\mu_{\tilde{A}}$ between 0 and 1 to be piecewise continuous. A fuzzy number $\tilde{A}$ with $\mu_{\tilde{A}}(x) = 1$ for some real numbers $x$ is called a normal fuzzy number. A fuzzy number $\tilde{A}$ is called an $LR$-type (normalized) fuzzy number if there are three real numbers $a \leq b \leq c$, an increasing function $L$ on the interval $[a, b]$ such that $L(a) = 0$ and $L(b) = 1$, and a decreasing function $R$ on the interval $[b, c]$ such that $R(b) = 1$ and $R(c) = 0$, satisfying for all $x$
New Similarity Measures Between Generalized Trapezoidal Fuzzy Numbers

\[ \mu_A(x) = \begin{cases} L(x), & \text{if } a \leq x \leq b \\ R(x), & \text{if } b \leq x \leq c \\ 0, & \text{elsewhere} \end{cases} \]

where \( b \) is called the mean value of \( A \) and \( \alpha = b - a \) and \( \beta = c - b \) are called the left and right spreads, respectively. \( \tilde{A} \) is denoted by \( \tilde{A} = (a, b, c)_L \). For an LR-type fuzzy number \( \tilde{A} = (a, b, c)_L \), if both of \( L \) and \( R \) are lines, then \( \tilde{A} \) is called a triangular fuzzy number and denoted by \( \tilde{A} = (a, b, c)_T \) with

\[ \mu_A(x) = \begin{cases} (x-a)/(b-a), & \text{if } a \leq x \leq b \\ (x-c)/(b-c), & \text{if } b \leq x \leq c \\ 0, & \text{elsewhere} \end{cases} \]

A fuzzy number \( \tilde{A} \) is called a LR-type (normalized) trapezoidal fuzzy number if there are real numbers \( a \leq b \leq c \leq d \) such that the membership \( \mu_A \) of \( \tilde{A} \) is

\[ \mu_A(x) = \begin{cases} L(x), & \text{if } a \leq x \leq b \\ 1, & \text{if } b \leq x \leq c \\ R(x), & \text{if } c \leq x \leq d \\ 0, & \text{elsewhere} \end{cases} \]

where \( \tilde{A} \) is denoted by \( \tilde{A} = (a, b, c, d)_L \). Furthermore, if we consider a weight \( w_\tilde{A} \) with \( 0 \leq w_\tilde{A} \leq 1 \) such that the fuzzy number \( \tilde{A} \) has its membership with

\[ \mu_A(x) = \begin{cases} w_\tilde{A}(x-a)/(b-a), & \text{if } a \leq x \leq b \\ w_\tilde{A}, & \text{if } b \leq x \leq c \\ w_\tilde{A}(x-d)/(c-d), & \text{if } c \leq x \leq d \\ 0, & \text{elsewhere} \end{cases} \]

The fuzzy number \( \tilde{A} \) is called a generalized trapezoidal fuzzy number (GTFN), and denoted by \( \tilde{A} = (a, b, c, d, w_\tilde{A}) \). When \( w_\tilde{A} = 1 \), \( \tilde{A} \) becomes to be a (normalized) trapezoidal fuzzy number. When \( w_\tilde{A} = 1 \) and \( b = c \), \( \tilde{A} \) becomes a triangular fuzzy number. LR-type fuzzy numbers are employed to present real numbers in a fuzzy environment, and GTFNs are utilized to present fuzzy intervals.

Without loss of generality, the left and right points of a GTFN \( \tilde{A} = (a, b, c, d, w_\tilde{A}) \) are reduced within the unit interval \([0,1]\), i.e., \( 0 \leq a \leq b \leq c \leq d \leq 1 \). Assume that there are two GTFNs \( \tilde{A} \) and \( \tilde{B} \) with \( \tilde{A} = (a_1, a_2, a_3, a_4, w_\tilde{A}) \) and \( \tilde{B} = (b_1, b_2, b_3, b_4, w_\tilde{B}) \) where \( 0 \leq a_1 \leq a_2 \leq a_3 \leq a_4 \leq 1 \) and \( 0 \leq b_1 \leq b_2 \leq b_3 \leq b_4 \leq 1 \). Chen\(^\text{12}\) presented a similarity measure \( S_c(\tilde{A}, \tilde{B}) \) between GTFNs \( \tilde{A} \) and \( \tilde{B} \) according to the geometric distance as follows:
$S_{c}(\tilde{A}, \tilde{B}) = 1 - \left(\sum_{i=1}^{4} |a_i - b_i| / 4\right)$.

Hsieh and Chen$^{13}$ presented a similarity measure $S_{hc}(\tilde{A}, \tilde{B})$ between GTFNs $\tilde{A}$ and $\tilde{B}$ using the graded mean integration representation distance as follows:

$$S_{hc}(\tilde{A}, \tilde{B}) = \frac{1}{1+d(\tilde{A}, \tilde{B})}$$

where $d(\tilde{A}, \tilde{B}) = |P(\tilde{A}) - P(\tilde{B})|$, $P(\tilde{A}) = \frac{a_1 + 2a_2 + 2a_3 + a_4}{6}$ and $P(\tilde{B}) = \frac{b_1 + 2b_2 + 2b_3 + b_4}{6}$.

Lee$^{14}$ presented a similarity measure $S_{c}(\tilde{A}, \tilde{B})$ between GTFNs $\tilde{A}$ and $\tilde{B}$ as follows:

$$S_{c}(\tilde{A}, \tilde{B}) = 1 - \frac{\|\tilde{A} - \tilde{B}\|_{\infty} \times 4^p}{\|U\|},$$

where $\|\tilde{A} - \tilde{B}\|_{\infty} = \max |a_i - b_i|$, $\|U\| = \max U - \min U$ and $P$ is a positive integer.

Chen and Chen$^{15}$ presented a similarity measure $S_{cc}(\tilde{A}, \tilde{B})$ between GTFNs $\tilde{A}$ and $\tilde{B}$ according to the combined concept of geometric distance and center of gravity (COG) distance of $\tilde{A}$ and $\tilde{B}$ as follows:

$$S_{cc}(\tilde{A}, \tilde{B}) = \left(1 - \sum_{i=1}^{4} |a_i - b_i| / 4\right)(1 - |x^*_{a} - x^*_{b}|) \frac{\min(y^*_{a}, y^*_{b})}{\max(y^*_{a}, y^*_{b})}$$

where:

$$y^*_{a} = \begin{cases} \frac{w_{i} \times (a_{i} - a_{i} + 2)}{6} & \text{if } a_{i} \neq a_{i} \\ \frac{w_{i}}{2} & \text{if } a_{i} = a_{i} \end{cases} \quad \text{and} \quad y^*_{b} = \begin{cases} \frac{w_{i} \times (b_{i} - b_{i} + 2)}{6} & \text{if } b_{i} \neq b_{i} \\ \frac{w_{i}}{2} & \text{if } b_{i} = b_{i} \end{cases}$$

and

$$x^*_{a} = \begin{cases} \frac{y_{a} (a_{i} + a_{i}) + (a_{i} + a_{i}) (w_{i} - y_{a})}{2w_{i}} & \text{if } w_{a} \neq 0 \\ \frac{a_{i} + a_{i}}{2} & \text{if } w_{a} = 0 \end{cases} \quad \text{and} \quad x^*_{b} = \begin{cases} \frac{y_{b} (b_{i} + b_{i}) + (b_{i} + b_{i}) (w_{i} - y_{b})}{2w_{i}} & \text{if } w_{b} \neq 0 \\ \frac{b_{i} + b_{i}}{2} & \text{if } w_{b} = 0 \end{cases}$$

and

$$B(S_{a}, S_{b}) = \begin{cases} 1 & \text{if } S_{a} + S_{b} > 0 \\ 0 & \text{if } S_{a} + S_{b} = 0 \end{cases} \quad \text{where} \quad S_{a} = a_{i} - a_{i}, S_{b} = b_{i} - b_{i}.$$

Wei and Chen$^{17}$ presented a similarity measure $S_{wc}(\tilde{A}, \tilde{B})$ between GTFNs $\tilde{A}$ and $\tilde{B}$ according to geometric distance and perimeters of $\tilde{A}$ and $\tilde{B}$ as follows:

$$S_{wc}(\tilde{A}, \tilde{B}) = \left(1 - \sum_{i=1}^{4} |a_i - b_i| / 4\right) \frac{\min(P(\tilde{A}), P(\tilde{B})) + \min(w_{a}, w_{b})}{\max(P(\tilde{A}), P(\tilde{B})) + \max(w_{a}, w_{b})}$$

where

$$P(\tilde{A}) = \sqrt{(a_{i} - a_{i})^2 + w_{a}^2} + a_{i} - a_{i} + \sqrt{(a_{i} - a_{i})^2 + w_{a}^2} + a_{i} - a_{i}$$

and
New Similarity Measures Between Generalized Trapezoidal Fuzzy Numbers

\[ P(\tilde{B}) = \sqrt{((b_1 - b_2)^2 + w_2^2) + b_3 - b_2 + \sqrt{((b_4 - b_3)^2 + w_3^2) + b_4 - b_1}}. \]

Xu et al.\textsuperscript{18} have recently modified Chen and Chen’s\textsuperscript{15} similarity measure by combining the concepts of geometric distance and center of gravity distance of GTFNs \( \tilde{A} \) and \( \tilde{B} \), and then proposed the similarity measure \( S_x(\tilde{A}, \tilde{B}) \) between GTFNs \( \tilde{A} \) and \( \tilde{B} \) as follows:

\[ S_x(\tilde{A}, \tilde{B}) = 1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{8} \cdot \frac{d(\hat{A}, \hat{B})}{2} \]
where \( d(\hat{A}, \hat{B}) = \frac{\sqrt{(x^*_{1} - y^*_{1})^2 +(y^*_{2} - y^*_{2})^2}}{\sqrt{1.25}} \).

In the next section, a new similarity measure between GTFNs \( \tilde{A} \) and \( \tilde{B} \) using the concept of Jaccard index is proposed.

### 3. A New Similarity Measure and Its Properties

In this section, a new similarity measure between GTFNs \( \tilde{A} \) and \( \tilde{B} \) using the Jaccard index (see Jaccard\textsuperscript{19} and Real and Vargas\textsuperscript{20}) is proposed. Let \( A \) and \( B \) be two sample sets. The Jaccard index \( J(A, B) \) is defined as the size of the intersection between \( \tilde{A} \) and \( \tilde{B} \) divided by the size of the union between \( \tilde{A} \) and \( \tilde{B} \). Let \( X_a \) and \( X_b \) be any two binary objects, each with \( n \) attributes, in which each attribute of the two objects can be either 0 or 1, then the binary Jaccard coefficient, \( J(X_a, X_b) \), is defined as the ratio of the number of shared attributes of \( X_a \) and \( X_b \) to the numbers possessed by \( X_a \) or \( X_b \) with

\[ J(X_a, X_b) = \frac{X_a \cdot X_b + X_a \cdot X_b - X_a \cdot X_b}{X_a \cdot X_a + X_b \cdot X_b}. \]

For example, if the two binary objects, \( X_a \) and \( X_b \), are given with binary indicator vectors \( X_a = (0,1,1,0) \) and \( X_b = (1,1,0,0) \), then the cardinality of their intersection is 1 and the cardinality of their union is 3, so the Jaccard index is 1/3. The Jaccard index has been widely employed to measure the similarity between two sample sets. Another good property of the Jaccard index is the simplicity of its calculation. To extend the binary Jaccard index to the fuzzy Jaccard index, a fuzzy object, \( \tilde{X} \), with \( n \) attributes, is defined as that each attribute of \( \tilde{X} \) is a fuzzy attribute by extending the value of each attribute with either 0 or 1 to the value between 0 and 1. Therefore, the fuzzy Jaccard index of the two fuzzy objects, \( \tilde{X}_a \) and \( \tilde{X}_b \), is defined as

\[ J(\tilde{X}_a, \tilde{X}_b) = \frac{\tilde{X}_a \cdot \tilde{X}_b}{\tilde{X}_a \cdot \tilde{X}_a + \tilde{X}_b \cdot \tilde{X}_b}. \]

According to this extended index, a new similarity measure between GTFNs \( \tilde{A} \) and \( \tilde{B} \) is constructed as follows.

Assume that the two GTFNs \( \tilde{A} \) and \( \tilde{B} \) are with \( \tilde{A} = (a_1, a_2, a_3, a_4, w_2) \) and \( \tilde{B} = (b_1, b_2, b_3, b_4, w_3) \), where \( 0 \leq a_1 \leq a_2 \leq a_3 \leq a_4 \leq 1 \) and \( 0 \leq b_1 \leq b_2 \leq b_3 \leq b_4 \leq 1 \), respectively. In general, a similarity measure between the two GTFNs \( \tilde{A} \) and \( \tilde{B} \) should
consider both criteria of their relative location and shape. For the relative location, the two terms with $m = \min \{ a_i, b_i \}$ and $M = \max \{ a_i, b_i \}$ are separately considered. For the shape, the weight ($w_{ij}$), the left and right spreads ($a_{3} - a_{4}$ and $a_{4} - a_{3}$), and the mean value ($a_{3} - a_{4}$) are included. Thus, the values under the term $m = \min \{ a_i, b_i \}$ are first constructed as follows:

$$
\begin{align*}
&x_i = a_i - m, i = 1, 2, 3, 4; \quad x_5 = w_{ij}; \quad x_6 = a_2 - a_1; \quad x_7 = a_3 - a_2; \quad x_8 = a_4 - a_3; \\
y_i = b_i - m, i = 1, 2, 3, 4; \quad y_5 = w_{ij}; \quad y_6 = b_2 - b_1; \quad y_7 = b_3 - b_2; \quad y_8 = b_4 - b_3.
\end{align*}
$$

According to the fuzzy Jaccard index concept, we consider the similarity $S_i(\tilde{A}, \tilde{B})$ between $\tilde{A}$ and $\tilde{B}$ under the term $m = \min \{ a_i, b_i \}$ as follows:

$$
S_i(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^{8} (x_i \times y_i)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i \times y_i)}
$$

Similarly, the values under the term $M = \max \{ a_i, b_i \}$ are constructed as follows:

$$
\begin{align*}
&x_{i+4} = M - x_i, i = 1, 2, 3, 4; \quad x_5' = x_5; \quad x_6' = x_6; \quad x_7' = x_7; \quad x_8' = x_8; \\
y_{i+4} = M - y_i, i = 1, 2, 3, 4; \quad y_5' = y_5; \quad y_6' = y_6; \quad y_7' = y_7; \quad y_8' = y_8.
\end{align*}
$$

Then the similarity $S_2(\tilde{A}, \tilde{B})$ between $\tilde{A}$ and $\tilde{B}$ under the term $M = \max \{ a_i, b_i \}$ are considered as follows:

$$
S_2(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^{8} (x_i' \times y_i')}{\sum_{i=1}^{8} x_i'^2 + \sum_{i=1}^{8} y_i'^2 - \sum_{i=1}^{8} (x_i' \times y_i')}
$$

Finally, a new similarity $S(\tilde{A}, \tilde{B})$ between $\tilde{A}$ and $\tilde{B}$ with an average of $S_i(\tilde{A}, \tilde{B})$ and $S_2(\tilde{A}, \tilde{B})$ is proposed as follows:

$$
S(\tilde{A}, \tilde{B}) = \frac{1}{2} \frac{\sum_{i=1}^{8} (x_i \times y_i)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i \times y_i)} + \frac{1}{2} \frac{\sum_{i=1}^{8} (x_i' \times y_i')}{\sum_{i=1}^{8} x_i'^2 + \sum_{i=1}^{8} y_i'^2 - \sum_{i=1}^{8} (x_i' \times y_i')}
$$

Thus, the proposed similarity measure $S(\tilde{A}, \tilde{B})$ has the following properties.

**Proposition 1.** $\tilde{A} = \tilde{B}$ if and only if $S(\tilde{A}, \tilde{B}) = 1$.

**Proof:** The part “only if” is claimed as follows:
Since $\tilde{A} = \tilde{B}$, then $x_i = y_i$ and $x_i^* = y_i^*$, $i = 1, \cdots, 8$. Therefore,

$$\frac{\sum_{i=1}^{8} (x_i \times y_i)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i \times y_i)} = \frac{\sum_{i=1}^{8} x_i^2}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} x_i^2 - \sum_{i=1}^{8} x_i^2} = 1$$

and

$$\frac{\sum_{i=1}^{8} (x_i^* \times y_i^*)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i^* \times y_i^*)} = \frac{\sum_{i=1}^{8} x_i^2}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} x_i^2} = 1.$$ 

Thus, $S(\tilde{A}, \tilde{B}) = \frac{1}{2} \sum_{i=1}^{8} (x_i \times y_i) + \frac{1}{2} \sum_{i=1}^{8} (x_i^* \times y_i^*) = 1$.

Next, the part “if” is claimed as follows: Since

$$\left( \sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i \times y_i) \right) - \left( \sum_{i=1}^{8} (x_i \times y_i) \right) = \sum_{i=1}^{8} (x_i^2 + y_i^2 - 2x_i y_i) = \sum_{i=1}^{8} (x_i - y_i)^2 \geq 0,$$

we have $\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i \times y_i) \geq \sum_{i=1}^{8} (x_i \times y_i) \geq 0$; that is,

$$0 \leq \frac{\sum_{i=1}^{8} (x_i \times y_i)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i \times y_i)} \leq 1.$$ 

Similarly, $0 \leq \frac{\sum_{i=1}^{8} (x_i^* \times y_i^*)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i^* \times y_i^*)} \leq 1$.

It is known that $S(\tilde{A}, \tilde{B}) = 1$ implies

$$\frac{1}{2} \sum_{i=1}^{8} (x_i \times y_i) + \frac{1}{2} \sum_{i=1}^{8} (x_i^* \times y_i^*) = 1.$$

Thus, we obtain

$$\frac{\sum_{i=1}^{8} (x_i \times y_i)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i \times y_i)} = \frac{\sum_{i=1}^{8} x_i^2}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} x_i^2 - \sum_{i=1}^{8} x_i^2} = 1$$

and

$$\frac{\sum_{i=1}^{8} (x_i^* \times y_i^*)}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} (x_i^* \times y_i^*)} = \frac{\sum_{i=1}^{8} x_i^2}{\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - \sum_{i=1}^{8} x_i^2} = 1.$$ 

This implies $\sum_{i=1}^{8} x_i^2 + \sum_{i=1}^{8} y_i^2 - 2 \sum_{i=1}^{8} (x_i \times y_i) = \sum_{i=1}^{8} (x_i - y_i)^2 = 0$ and $\sum_{i=1}^{8} (x_i^* - y_i^*)^2 = 0$.

Therefore, we have that $x_i = y_i$ and $x_i^* = y_i^*$, $i = 1, \cdots, 8$. Hence, $\tilde{A} = \tilde{B}$. □
Proposition 2. \( S(\widetilde{A}, \widetilde{B}) = S(\widetilde{B}, \widetilde{A}) \).

Proof: Since \( \sum_{i=1}^{8} (x_i \times y_i) = \sum_{i=1}^{8} (y_i \times x_i) \) and \( \sum_{i=1}^{8} (x_i^* \times y_i^*) = \sum_{i=1}^{8} (y_i^* \times x_i^*) \), it is easy to show that \( S(\widetilde{A}, \widetilde{B}) = S(\widetilde{B}, \widetilde{A}) \).

Proposition 3. For any generalized trapezoidal fuzzy numbers \( \widetilde{A} = (a_1, a_2, a_3, a_4, w_{\widetilde{A}}) \) and \( \widetilde{B} = (b_1, b_2, b_3, b_4, w_{\widetilde{B}}) \). We have that \( S(\widetilde{A}, \widetilde{B}) = S(\widetilde{A}_i, \widetilde{B}_i) \) for \( \widetilde{A}_i = \theta \widetilde{A} = (\theta a_1, \theta a_2, \theta a_3, \theta a_4, \theta w_{\widetilde{A}}) \) and \( \widetilde{B}_i = \theta \widetilde{B} = (\theta b_1, \theta b_2, \theta b_3, \theta b_4, \theta w_{\widetilde{B}}) \), \( \theta > 0 \).

Proof: Since \( \widetilde{A}_i = \theta \widetilde{A} \) and \( \widetilde{B}_i = \theta \widetilde{B} \), we have

\[
S(\widetilde{A}_i, \widetilde{B}_i) = \frac{1}{2} \sum_{i=1}^{8} (\theta x_i \times \theta y_i) + \frac{1}{2} \sum_{i=1}^{8} (\theta x_i^* \times \theta y_i^*) = \frac{1}{2} \sum_{i=1}^{8} (x_i \times y_i) + \frac{1}{2} \sum_{i=1}^{8} (x_i^* \times y_i^*) = S(\widetilde{A}, \widetilde{B})
\]

That is, \( S(\widetilde{A}_i, \widetilde{B}_i) = S(\widetilde{A}, \widetilde{B}) \). \( \square \)

Note that the new similarity \( S(\widetilde{A}, \widetilde{B}) \) between \( \widetilde{A} \) and \( \widetilde{B} \) may be extended with a 0 \( \leq \alpha \leq 1 \) convex combination of \( S_1(\widetilde{A}, \widetilde{B}) \) and \( S_2(\widetilde{A}, \widetilde{B}) \) as follows:

\[
S_\alpha(\widetilde{A}, \widetilde{B}) = \frac{\alpha}{8} \sum_{i=1}^{8} (x_i \times y_i) + \frac{1-\alpha}{8} \sum_{i=1}^{8} (x_i^* \times y_i^*)
\]

\( S_\alpha(\widetilde{A}, \widetilde{B}) \) can be similarly claimed to satisfy Propositions 1, 2 and 3. In fact, the weighted parameter \( \alpha \) in \( S_\alpha(\widetilde{A}, \widetilde{B}) \) does not have much effect on most cases of \( \widetilde{A} \) and \( \widetilde{B} \), so \( \alpha = 1/2 \) is generally chosen.

Example 1. Let the three trapezoidal fuzzy numbers be \( \widetilde{A} = (0.15, 0.25, 0.25, 0.35, 1) \), \( \widetilde{B} = (0.2, 0.3, 0.3, 0.4, 1) \) and \( \widetilde{C} = (0.2, 0.25, 0.25, 0.3, 1) \). Then \( S(\widetilde{A}, \widetilde{B}) = 0.991 \), \( S(\widetilde{B}, \widetilde{C}) = 0.9815 \) and \( S(\widetilde{A}, \widetilde{C}) = 0.991 \). That is, \( \widetilde{B} \) and \( \widetilde{C} \) are the least similar among \( \widetilde{A}, \widetilde{B} \) and \( \widetilde{C} \). This is consistent with the intuition for the three trapezoidal fuzzy numbers \( \widetilde{A}, \widetilde{B} \) and \( \widetilde{C} \). On the other hand, we find that \( S_\alpha(\widetilde{A}, \widetilde{B}) = 0.991 \) for all \( \alpha \).
New Similarity Measures Between Generalized Trapezoidal Fuzzy Numbers

\[ S_{\alpha}(\tilde{B}, \tilde{C}) = \alpha 0.981 + (1 - \alpha) 0.982 \]
\[ S_{\alpha}(\tilde{A}, \tilde{C}) = 0.991 \] for all \( \alpha \). That is, \( S_{\alpha}(\tilde{A}, \tilde{B}) \) and \( S_{\alpha}(\tilde{A}, \tilde{C}) \) are independent of \( \alpha \), but \( S_{\alpha}(\tilde{B}, \tilde{C}) \) is dependent on \( \alpha \). However, they also have the same result that \( \tilde{B} \) and \( \tilde{C} \) are the least similar among \( \tilde{A}, \tilde{B} \) and \( \tilde{C} \).

4. Examples and Comparisons

In this section, numerical comparisons of the proposed similarity measures with some existing ones developed by Chen,12 Hsieh and Chen,13 Lee,14 Chen and Chen,15 Wei and Chen17 and Xu et al.18 are made.

Example 2. There are 15 sets of fuzzy numbers shown in Fig. 1, which was used in Wei and Chen.17 The results obtained by the proposed and existing similarity measures are shown in Table 1. From Table 1, we can see the drawbacks of these existing similarity measures Chen,12 Hsieh and Chen,13 Lee14 and Chen and Chen15 that had been described in Wei and Chen.17 Furthermore, we also add the results of Xu et al.18 in Table 1.

Analysis of the results shown in Table 1 is summarized as follows.

(1) As seen in Sets 1 and 5, \( \tilde{A} \) and \( \tilde{B} \) are two different GTFNs. However, the degree of similarity between \( \tilde{A} \) and \( \tilde{B} \) obtained by Hsieh and Chen’s measure\(^{13}\) for Set 1 is 1, while that obtained by the measures of Chen,12 Hsieh and Chen13 and Lee\(^{14}\) for Set 5 is also 1, both of which are incorrect.

(2) Comparing Set 3 with Set 4 shows that the two different GTFNs in Set 4 should be more similar than those in Set 3. However, the measures of Chen,12 Hsieh and Chen13 and Lee14 all gave the same degree of similarity, which is again incorrect.

(3) Sets 6 and 7 reveal that Lee’s measure\(^{14}\) yielded incorrect results. It could not calculate the degree of similarity for two identical real values in Set 6, and obtained zero degree of similarity for the two different real values in Set 7.

(4) Comparing Set 8 with Set 9 shows that they are two different sets of GTFNs. However, as seen in Table 1, the measure of Chen\(^{12}\) and Hsieh and Chen\(^{13}\) gave the same degree of similarity, which is incorrect.

(5) Comparing Set 9 with Set 10 reveals that the two different GTFNs in Set 9 have exactly the same shape, but the two different GTFNs in Set 10 have the same mean location at 0.4. With both criteria of shape and location similarities taken into consideration, the two different GTFNs in Set 9 are more similar than those in Set 10. However, Table 1 shows that all measures, except that of Wei and Chen\(^{17}\) and the proposed similarity measure, provided incorrect results. In fact, the proposed similarity \( S(\tilde{A}, \tilde{B}) \) for the two different GTFNs \( \tilde{A} \) and \( \tilde{B} \) had considered weight as well as left and right spreads, which are connected with shape and location of the GTFNs. In this way, it took both shape and location similarities and yielded the result that the two different GTFNs in Set 9 are more similar than those in Set 10.

(6) Comparing Set 14 with Set 15 shows that the two different GTFNs in Set 14 are more similar than those in Set 15. However, as seen in Table 1, Chen and Chen’s method\(^{15}\) gave incorrect results.
Overall, only Wei-Chen’s measure\(^{17}\) and the proposed similarity measure gave all correct results for Sets 1–15 in this example. Furthermore, consider \(S_\alpha(\tilde{A}, \tilde{B})\) with a weight \(\alpha\), it is found that all \(S_\alpha(\tilde{A}, \tilde{B})\) values for Sets 1–15 are independent of \(\alpha\), except \(S_\alpha(\tilde{A}, \tilde{B}) = \alpha 0.714 + (1-\alpha)0.780\) for Set 3, \(S_\alpha(\tilde{A}, \tilde{B}) = \alpha 0.931 + (1-\alpha)0.926\) for Set 8, and \(S_\alpha(\tilde{A}, \tilde{B}) = \alpha 0.903 + (1-\alpha)0.927\) for Set 12. However, all comparison results from \(S_\alpha(\tilde{A}, \tilde{B})\) are exactly the same as those from \(S(\tilde{A}, \tilde{B})\).

![Fig. 1. Fifteen sets of generalized fuzzy numbers for Example 1.](image-url)
Table 1. Comparison of the proposed similarity $S(\tilde{A},\tilde{B})$ with the existing methods.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>0.9617</td>
<td>1</td>
<td>0.975</td>
<td>0.8357</td>
<td>0.950</td>
<td>0.9627</td>
<td>0.985</td>
</tr>
<tr>
<td>Set 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Set 3</td>
<td>0.5</td>
<td>0.7692</td>
<td>0.7</td>
<td>0.42</td>
<td>0.682</td>
<td>0.714</td>
<td>0.747</td>
</tr>
<tr>
<td>Set 4</td>
<td>0.5</td>
<td>0.7692</td>
<td>0.7</td>
<td>0.49</td>
<td>0.7</td>
<td>0.799</td>
<td>0.795</td>
</tr>
<tr>
<td>Set 5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.8248</td>
<td>0.955</td>
<td>0.96</td>
</tr>
<tr>
<td>Set 6</td>
<td>*</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Set 7</td>
<td>0</td>
<td>0.9091</td>
<td>0.9</td>
<td>0.81</td>
<td>0.9</td>
<td>0.905</td>
<td>0.962</td>
</tr>
<tr>
<td>Set 8</td>
<td>0.5</td>
<td>0.9091</td>
<td>0.9</td>
<td>0.54</td>
<td>0.8411</td>
<td>0.866</td>
<td>0.929</td>
</tr>
<tr>
<td>Set 9</td>
<td>0.6667</td>
<td>0.9091</td>
<td>0.9</td>
<td>0.81</td>
<td>0.9</td>
<td>0.905</td>
<td>0.966</td>
</tr>
<tr>
<td>Set 10</td>
<td>0.8333</td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7833</td>
<td>0.95</td>
<td>0.902</td>
</tr>
<tr>
<td>Set 11</td>
<td>0.75</td>
<td>1</td>
<td>0.9</td>
<td>0.72</td>
<td>0.8003</td>
<td>0.913</td>
<td>0.915</td>
</tr>
<tr>
<td>Set 12</td>
<td>0.8</td>
<td>0.9375</td>
<td>0.9</td>
<td>0.8325</td>
<td>0.8289</td>
<td>0.89</td>
<td>0.915</td>
</tr>
<tr>
<td>Set 13</td>
<td>0.75</td>
<td>0.9091</td>
<td>0.9</td>
<td>0.81</td>
<td>0.9</td>
<td>0.905</td>
<td>0.969</td>
</tr>
<tr>
<td>Set 14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.7</td>
<td>0.7209</td>
<td>0.992</td>
<td>0.87</td>
</tr>
<tr>
<td>Set 15</td>
<td>0.75</td>
<td>1</td>
<td>0.95</td>
<td>0.9048</td>
<td>0.6215</td>
<td>0.926</td>
<td>0.851</td>
</tr>
</tbody>
</table>

**Example 3.** Assume that there are the two sets of GTFNs, where Set 16 is with $\{\tilde{A} = (0.25,0.25,0.35,0.35,1.0), \tilde{B} = (0.45,0.45,0.75,0.75,1.0)\}$, and Set 17 is with $\{\tilde{A}_1 = (0.25,0.25,0.35,0.35,1.0), \tilde{B}_1 = (0.55,0.55,0.65,0.65,1.0)\}$ as shown in Fig. 2. Hence, $S_c(\tilde{A},\tilde{B}) = 0.769$, $S_{cc}(\tilde{A},\tilde{B}) = 0.75$, $S_{cc}(\tilde{A}_1,\tilde{B}_1) = 0.49$, and $S(\tilde{A},\tilde{B}) = 0.72 < S(\tilde{A}_1,\tilde{B}_1) = 0.752$. As can be seen, the measure of Chen, Hsieh and Chen, Chen and Chen and Xu et al. yielded incorrect results, while that of Lee, Wei and Chen and the proposed similarity measure gave correct results.

**Example 4.** Assume that there are the two sets of GTFNs, where Set 18 is with $\{\tilde{A} = (0.1,0.2,0.3,0.4,1), \tilde{B} = (0.4,0.5,0.6,0.7,1.0)\}$, and Set 19 is with $\{\tilde{A}_1 = (0.1,0.2,0.3,0.4,0.5), \tilde{B}_1 = (0.4,0.5,0.6,0.7,0.5)\}$ as shown in Fig. 3. The two GTFNs in Set 18 have the same base (width) (i.e. exactly the same location) as the two GTFNs in Set 19, but only half the height (i.e., of different scales). Overall, the similarity between the two GTFNs in Set 18 should be larger than that between the two...
GTFNs in Set 19. The calculations yielded \( S_C(\tilde{A}, \tilde{B}) = S_C(\tilde{A}_1, \tilde{B}_1) = 0.7 \), \( S_{hw}(\tilde{A}, \tilde{B}) = S_{hw}(\tilde{A}_1, \tilde{B}_1) = 0.7692 \), \( S_h(\tilde{A}, \tilde{B}) = S_h(\tilde{A}_1, \tilde{B}_1) = 0.5 \), \( S_{hc}(\tilde{A}, \tilde{B}) = S_{hc}(\tilde{A}_1, \tilde{B}_1) = 0.49 \), \( S_{ws}(\tilde{A}, \tilde{B}) = S_{ws}(\tilde{A}_1, \tilde{B}_1) = 0.7 \), \( S_w(\tilde{A}, \tilde{B}) = S_w(\tilde{A}_1, \tilde{B}_1) = 0.72 \), \( S(\tilde{A}, \tilde{B}) = 0.79 > S(\tilde{A}_1, \tilde{B}_1) = 0.63 \).

As can be seen, the measures of Chen, Hsieh and Chen, Lee, Chen and Chen, Wei and Chen, and Xu et al. got equal similarity between the two sets, which is incorrect. In this example, only the proposed similarity measure yielded correct results.

\[
\begin{align*}
&\tilde{A} = (0.25, 0.45, 0.55, 0.75, 1.0) \\
&\tilde{B} = (0.45, 0.45, 0.75, 0.75, 1.0)
\end{align*}
\]

\[
\begin{align*}
&\tilde{A}_1 = (0.05, 0.1, 0.15, 0.2, 0.5) \\
&\tilde{B}_1 = (0.2, 0.25, 0.3, 0.35, 0.5)
\end{align*}
\]

Fig. 2. Two sets of generalized fuzzy numbers for Example 3.

Fig. 3. Two sets of generalized fuzzy numbers for Example 4.

**Example 5.** Assume that there are two sets of GTFNs, where Set 20 is with \( \{\tilde{A} = (0.1, 0.2, 0.3, 0.4, 1), \tilde{B} = (0.4, 0.5, 0.6, 0.7, 1.0)\} \), and Set 21 is with \( \{\tilde{A}_1 = (0.05, 0.1, 0.15, 0.2, 0.5), \tilde{B}_2 = (0.2, 0.25, 0.3, 0.35, 0.5)\} \) as shown in Fig. 4. Comparing Set 20 with Set 21 shows that the two GTFNs in Set 20 are just double in scale of the two GTFNs in Set 21. In general, a good similarity should be scale-free. According to the calculations, \( S_C(\tilde{A}, \tilde{B}) = 0.7 < S_C(\tilde{A}_1, \tilde{B}_1) = 0.85 \), \( S_L(\tilde{A}, \tilde{B}) = 0.5 < S_L(\tilde{A}_1, \tilde{B}_1) = 0.81 \), \( S_{hc}(\tilde{A}, \tilde{B}) = 0.7692 < S_{hc}(\tilde{A}_1, \tilde{B}_1) = 0.8 \), \( S_{cc}(\tilde{A}, \tilde{B}) = 0.49 < S_{cc}(\tilde{A}_1, \tilde{B}_1) = 0.723 \), \( S_w(\tilde{A}, \tilde{B}) = 0.70 < S_w(\tilde{A}_1, \tilde{B}_1) = 0.85 \), \( S_w(\tilde{A}, \tilde{B}) = 0.799 < S_w(\tilde{A}_1, \tilde{B}_1) = 0.919 \), and \( S(\tilde{A}, \tilde{B}) = S(\tilde{A}_1, \tilde{B}_1) = 0.79 \). As can be
seen, the measure of Chen, Hsieh and Chen, Lee, Chen and Chen, Wei and Chen and Xu et al. are scale-dependent, but our result of $S(\tilde{A}, \tilde{B}) = S(\tilde{A}_r, \tilde{B}_r) = 0.79$ obtained by the proposed similarity measures shows that it is scale-free.

5. Conclusions

In this paper, we develop new similarity measures between GTFNs using the Jaccard index. The Jaccard index is a useful tool for comparing the similarity and diversity of sample sets; hence, similarity measures between GTFNs developed using the Jaccard index can reveal the degree of similarity between GTFNs. Examples are provided to illustrate that similarity measures developed using the Jaccard index can deal with problems more reasonably than most existing similarity measures for GTFNs.

Acknowledgements

The authors are grateful to the anonymous referees for their constructive comments and suggestions to improve the presentation of the paper. This work was supported in part by the National Science Council of Taiwan, under Grant NSC101-2118-M-033-002-MY2.

References


