An Intelligent Similarity Model between Generalized Trapezoidal Fuzzy Numbers in Large Scale

Mohamedou Cheikh Tourad and Abdelmounaim Abdali
Applied Mathematics and Computer Science Laboratory, Cadi Ayyad University, Marrakech, Morocco

Abstract
The rapid expansion of data published on the web has given rise to the similarity problem on a large scale, a very important subject for scientific research in the field of computer science. Several methods have been developed for this. In this paper, we propose the first mathematical model to find the similarity value between generalized trapezoidal fuzzy numbers (GTFNs). This model employs fuzzy inference systems to find the value of an effective weighting, the weights to be associated to different kinds of methods that can handle an important scale of the data. This model will allow us to develop intelligent systems. A comparative study based on 21 sets of GTFNs has been carried out to demonstrate the difference between our approach and existing methods. This study shows that our model is more reasonable than existing methods.

Keywords: Cosine coefficient, Jaccard index, GTFNs, FIS, Similarity, Large scale

1. Introduction
The concept of fuzzy logic, proposed in 1965 by Zadeh [1], has been used to manage a kind of probability. To employ this concept, the generalized trapezoidal fuzzy numbers (GTFNs) are most the most popular in practice. In the literature, several similarity methods between GTFNs have been introduced (e.g., [2–10]). But, these existing methods of similarity measures have many weaknesses. In many situations, such methods cannot appropriately find the similarity between two GTFNs. In the present study, a novel mathematical model for a fuzzy-number similarity method between GTFNs has been created, based on the weights associated to each similarity measure. This model uses the cosine coefficient and the Jaccard Index. Additionally, we describe and provide three characteristics of the proposed model. A comparative study has been carried out, based on 21 sets of GTFNs, to show that the model can surmount the limitations of the existing measures.

The remainder of this article is structured as follows. Section 2 presents a summary of the fundamental notions of existing methods. Section 3 introduces the novel approach, presenting the similarity methods employed and finding the weights associated to each similarity method. Many properties are suggested and proven. Section 4 compares it with the existing similarity methods. We give our conclusions in Section 5.

2. Related Work
The notion of a GFN T is presented by Chen [11, 12] as follows: $T = \{t_1, t_2, t_3, t_4, w_T\}$, where
where
\[ y^*_F = \begin{cases} \frac{w_F}{6} \times (2 + \frac{t_4 - t_1}{t_4 - t_1}) & \text{if } t_4 \neq t_1, \\ \frac{w_F}{2} & \text{if } t_4 = t_1, \end{cases} \]
\[ y^*_H = \begin{cases} \frac{w_H}{6} \times (2 + \frac{h_3 - h_4}{h_3 - h_4}) & \text{if } h_4 \neq h_1, \\ \frac{w_H}{2} & \text{if } h_4 = h_1, \end{cases} \]
\[ x^*_F = \frac{((w_F \times (t_4 + t_2) + (t_4 + t_1) \times (w_F - w_H))}{2w_F} \text{ if } w_F \neq 0, \]
\[ x^*_H = \frac{((w_H \times (h_3 + h_2)) + (h_4 + h_1) \times (w_H - w_H))}{2w_H} \text{ if } w_H \neq 0, \]
\[ B(S_T, S_H) = \begin{cases} 1 & \text{if } S_T + S_H > 0, \\ 0 & \text{if } S_T + S_H = 0, \end{cases} \]
where
\[ S_T = t_4 - t_1, \quad S_H = h_4 - h_1. \]
Wei and Chen [7] proposed a new approach in order to solve the similarity problems of GFNs. It is based on the concepts of geometric distance, the perimeter and the height of GFNs. The authors used this approach to answer the similarity problems of fuzzy risk analysis.

\[ S_{WC}(T, H) = (1 - \left( \sum_{k=1}^{4} \frac{t_k - h_k}{4} \right)) \times \left( \frac{\min(P(T), P(H)) + \min(w_F, w_H)}{\max(P(T), P(H)) + \max(w_F, w_H)} \right), \]

where
\[ P(T) = \sqrt{(t_4 - t_2)^2 + (w_F)^2 + (t_3 - t_2)} + \sqrt{(t_4 - t_3)^2 + (w_F)^2} + (t_4 - t_1), \]
\[ P(H) = \sqrt{(h_3 - h_2)^2 + (w_H)^2} + (h_3 - h_2) + \sqrt{(h_4 - h_3)^2 + (w_H)^2} + (h_4 - h_1). \]

Recently, Xu et al. [8] changed the similarity method SCC proposed by Chen and Chen [5], introducing the new similarity \( S_X \). The authors used this method to answer the problems of fuzzy risk analysis.

\[ S_X(T, H) = 1 - \left( \sum_{k=1}^{4} \frac{t_k - h_k}{8} - \frac{d(T, H)}{2} \right), \]

where \[ d(T, H) = \sqrt{(x^*_T - x^*_H)^2 + (y^*_T - y^*_H)^2} \].

Chen [9] introduced a new proposed method to find a solution
to the similarity problems of GFNs. This method is based on the geometric mean operator. The author used this approach to clear the problems of fuzzy-number used in the information retrieval.

\[
S_{j,i}(\tilde{T}, \tilde{H}) = \left[\prod_{k=1}^{4} (2 - \|t_k - h_k\|)\right]^{\frac{1}{2}} - 1 \times \min(y_{\tilde{T}}^*, y_{\tilde{H}}^*) \div \max(y_{\tilde{T}}^*, y_{\tilde{H}}^*) \left(7\right)
\]

where \(y_{\tilde{T}}^*\) and \(y_{\tilde{H}}^*\) are calculated as in Chen and Chen [5].

3. Approach

We will now present a novel mathematical model, MCESTA (Mohamedou Cheikh Elghotob Cheikh Saad bouch Cheikh Tourad Abass), Abdelmounaim Abdali, a large scale similarity method between GTFNs, call them \(\tilde{T}\) and \(\tilde{H}\). It is a kind of hybrid of the similarity measures \(S_k\) and \(S_k^q\), where each measure is associated with an importance weight (\(\alpha_k\) and \(\beta_q\)). MCESTA is an efficient development of the formula FBSHSM=\(\sum_{k=1}^{n} \alpha_k \times S_k\) proposed by Gupta et al. [13] to calculate similarity in IR (information retrieval). MCESTA employs a fuzzy inference system (FIS) to find the value of the effective weighting. This hybrid similarity measure is calculated as follows:

\[
MCESTA(\tilde{T}, \tilde{H}) = \left\{\begin{array}{ll}
\sum_{k=1}^{n} \alpha_k \times S_k(\tilde{T}, \tilde{H}), & S_k(\tilde{T}, \tilde{H}) = \sum_{q=1}^{m_k} \beta_q \times S_k^q(\tilde{T}, \tilde{H}), \\
\end{array}\right.  
\]

where \(\sum_{k=1}^{n} \alpha_k \leq 1\) and \(\sum_{q=1}^{m_k} \beta_q \leq 1\).

Our model will be as follows:

\[
MCESTA(\tilde{T}, \tilde{H}) = \sum_{K=1}^{n} \alpha_k \times \sum_{q=1}^{m_k} \beta_q S_k^q(\tilde{T}, \tilde{H}), \left(8\right)
\]

where \(S_k\) is a similarity method between \(\tilde{T}\) and \(\tilde{H}\), \(S_k^q\) is a similarity sub-measure between \(\tilde{T}\) and \(\tilde{H}\) that should take into account the two criteria of their relative location and size. This measure is calculated to determine a reasonable value of similarity \(S_k\). \(\alpha_k\) is an importance weight associated with \(S_k\) and \(\beta_q\) is an importance weight associated with \(S_k^q\). The number \(n\) defines the number of similarity methods \(S_k\) used to calculate the model. The number \(m_k\) defines the number of similarity sub-measures \(S_k^q\) employed to calculate \(S_k\).

3.1 Similarity Measures \(S_k\) and \(S_k^q\)

The descriptions of the similarity methods employed, the cosine coefficient and the Jaccard Index [14], are given below. This choice of similarity measures will be validated in Section 4. We have \(n = 2\) and \(S_1 = \cos\) and \(S_2 = Jaccard\) and we choose \(m_1 = 1\) and \(m_2 = 2\), which implies, in (9),

\[
MCESTA(\tilde{T}, \tilde{H}) = \sum_{k=1}^{2} \alpha_k \times \sum_{q=1}^{m_k} \beta_q S_k^q(\tilde{T}, \tilde{H}),
\]

for each \(k\) we provide a corresponding \(m_k\).

\[
MCESTA(\tilde{T}, \tilde{H}) = \alpha_1 \times \sum_{q=1}^{m_1} \beta_q S_1^q(\tilde{T}, \tilde{H}) + \alpha_2 \times \sum_{q=1}^{m_2} \beta_q S_2^q(\tilde{T}, \tilde{H}),
\]

where \(S_1^1 = \cos\), \(S_1^2 = Jaccard\) and \(S_2^2 = Jaccard\). To facilitate the calculation of this approach, one needs to make the following changes of variable: \(C = \alpha_1 \times \beta_1\), \(J_1 = \alpha_2 \times \beta_1\) and \(J_2 = \alpha_2 \times \beta_2\). So this equation will become

\[
MCESTA(\tilde{T}, \tilde{H}) = C \times \cos(\tilde{T}, \tilde{H}) + J_1 \times Jaccard_1(\tilde{T}, \tilde{H}) + J_2 \times Jaccard_2(\tilde{T}, \tilde{H}). \left(9\right)
\]

3.1.1 The cosine coefficient \(\cos(\tilde{T}, \tilde{H})\)

The cosine coefficient calculates the similarity between vectors in an easy and more intelligent way: it is by the determination of the direction. \(\cos(\tilde{T}, \tilde{H})\) is calculated as follows:

\[
x_k = t_k,
\]

where \(k = \{1, 2, 3, 4\}\), \(x_5 = w_T\), \(x_6 = t_2 - t_1\), \(x_7 = t_3 - t_2\), \(x_8 = t_4 - t_3\). Also, \(y_k = h_k\) where \(k = \{1, 2, 3, 4\}\), \(y_5 = w_H\), \(y_6 = h_2 - h_1\), \(y_7 = h_3 - h_2\), \(y_8 = h_4 - h_3\).

\[
\cos(\tilde{T}, \tilde{H}) = \frac{\sum_{k=1}^{8} x_k \times y_k}{(\sum_{k=1}^{8} x_k^2) \times (\sum_{k=1}^{8} y_k^2)} \left(10\right).
\]

3.1.2 The Jaccard Index

The Jaccard Index, \(Jaccard(\tilde{T}, \tilde{H})\) [10], measures the similarity as follows: the size of the intersection is divided by the size of the union.
3.2 The Methods to Define the Weights $C$, $J_1$ and $J_2$

Our choice is based on an FIS, such systems are recognized and have been used in many different fields \cite{15,16}. To calculate the weights mentioned in Eq. \ref{eq:weights} we use the Mamdani-type FIS mentioned in \cite{13,17,19} adapted to our model. Figure \ref{fig:mamdani} shows the Mamdani for MCESTA.

\begin{equation}
Jaccard_1(T, H) = \frac{\sum_{k=1}^{8} x_k \times y_k}{(\sum_{k=1}^{8} x_k^2) + (\sum_{k=1}^{8} y_k^2) - \sum_{k=1}^{8} x_k \times y_k}.
\end{equation}

\begin{equation}
Jaccard_2(T, H) = \frac{\sum_{k=1}^{8} x_k^* \times y_k^*}{\sum_{k=1}^{8} (x_k^*)^2 + \sum_{k=1}^{8} (y_k^*)^2 - \sum_{i=k}^{8} x_k \times y_k^*}.
\end{equation}

$Jaccard_1$ is calculated as follows: $x_k = t_k - m$, where $[k = \{1, 2, 3, 4\}]$, $x_5 = w_T$, $x_6 = t_2 - t_1$, $x_7 = t_3 - t_2$, $x_8 = t_4 - t_3$, and $y_k = h_k - m$, where $[k = \{1, 2, 3, 4\}]$, $y_5 = w_H$, $y_6 = h_2 - h_1$, $y_7 = h_3 - h_2$, $y_8 = h_4 - h_3$, where $m = \min(t_1, h_1)$.

In Table \ref{tab:fuzzyrules} there is presented the FIS fuzzy rule for the MCESTA, which is a set of semantic declarations that define how the FIS-Mamdani must carry out its decision making for input state (Cosine, Jaccard$_1$, and Jaccard$_2$) or controlling an output ($C$, $J_1$ and $J_2$). This table uses the values of the functions presented in Figure \ref{fig:jaccard}.

In Figure \ref{fig:rule} the FIS-diagram for the MCESTA defined as a procedure for developing the association relation (from a given value to an output value) based on the concepts of fuzzy logic. This figure shows the values which are the most important for the weights following $C$, $J_1$ and $J_2$.

Figure \ref{fig:rule} to Figure \ref{fig:rule9} show the evolution of the FIS-Rule surface diagram for the weights ($C$, $J_1$ and $J_2$) in MCESTA.

3.3 Properties

...
Validating their proposed similarity method. We have that

### 3.3.1 Identical

\[ \tilde{T} = \tilde{H} \Leftrightarrow MCESTA(\tilde{T}, \tilde{H}) = 1. \]

### 3.3.2 Symmetric relation

\[ MCESTA(\tilde{T}, \tilde{H}) = MCESTA(\tilde{H}, \tilde{T}). \]

### 3.3.3 Scale-free

This property has been used by Hwang and Yang [10] for validating their proposed similarity method. We have that

\[ MCESTA(\tilde{T}_1, \tilde{H}_1) = MCESTA(\tilde{T}, \tilde{H}) \]

for \( \tilde{T}_1 = \alpha \times \tilde{T} = (\alpha \times t_1, \alpha \times t_2, \alpha \times t_3, \alpha \times t_4, \alpha \times w_{\tilde{T}}) \) and \( \tilde{H}_1 = \alpha \times \tilde{H} = (\alpha \times h_1, \alpha \times h_2, \alpha \times h_3, \alpha \times h_4, \alpha \times w_{\tilde{H}}) \), where \( \alpha \geq 0 \). Normally, the best similarity should be scale free. This is a very significant property for the large scale. According to this result, we deduce that a very large data \( Set_1 = (\tilde{T}_1, \tilde{H}_1) \) can be analyzed using another batch of data of small size \( Set = (\tilde{T}, \tilde{H}) \). But according to the scale relation \( \tilde{T}_1 = \alpha \times \tilde{T} \) and \( \tilde{H}_1 = \alpha \times \tilde{H} \), the interpretation and the analysis of \( Set \) gives a global idea of the interpretation of \( Set_1 \), which is important when it can be difficult to treat everything in the context of big data or the machines don’t have sufficient capacity. This property is very important for large scale studies.

### Table 1. FIS fuzzy rule for MCESTA

| Rule | Description |
|------|-------------|
| IF (Jaccard1 is L) and (Jaccard2 is L) and (Cos is L) THEN (\( \tilde{J}_1 \) is L) (\( \tilde{J}_2 \) is L) (C is L) (1). |
| IF (Jaccard1 is L) and (Jaccard2 is M) and (Cos is L) THEN (\( \tilde{J}_1 \) is L) (\( \tilde{J}_2 \) is M) (C is L) (1). |
| IF (Jaccard1 is L) and (Jaccard2 is H) and (Cos is L) THEN (\( \tilde{J}_1 \) is L) (\( \tilde{J}_2 \) is H) (C is L) (1). |
| IF (Jaccard1 is M) and (Jaccard2 is L) and (Cos is L) THEN (\( \tilde{J}_1 \) is M) (\( \tilde{J}_2 \) is L) (C is L) (1). |
| IF (Jaccard1 is M) and (Jaccard2 is M) and (Cos is L) THEN (\( \tilde{J}_1 \) is M) (\( \tilde{J}_2 \) is M) (C is L) (1). |
| IF (Jaccard1 is M) and (Jaccard2 is H) and (Cos is L) THEN (\( \tilde{J}_1 \) is M) (\( \tilde{J}_2 \) is H) (C is L) (1). |
| IF (Jaccard1 is H) and (Jaccard2 is L) and (Cos is L) THEN (\( \tilde{J}_1 \) is H) (\( \tilde{J}_2 \) is L) (C is L) (1). |
| IF (Jaccard1 is H) and (Jaccard2 is M) and (Cos is L) THEN (\( \tilde{J}_1 \) is H) (\( \tilde{J}_2 \) is M) (C is L) (1). |
| IF (Jaccard1 is H) and (Jaccard2 is H) and (Cos is L) THEN (\( \tilde{J}_1 \) is H) (\( \tilde{J}_2 \) is H) (C is L) (1). |

L, Low; M, Medium; H, High.
3.4 Proof of Properties

To see that our model verifies the three properties, it is enough that every similarity (Cosine, Jaccard$_1$ and Jaccard$_2$) used in this model satisfies these properties. For the sub-measures Jaccard$_1$ and Jaccard$_2$, this is proved by Hwang and Yang [10]. It remains to treat Cosine.

3.4.1 MCESTA is identical

\[ \tilde{T} = \tilde{H} \iff \cos(\tilde{T}, \tilde{H}) = 1. \]

(i) The \( \Rightarrow \) is proved by observing that since \( \tilde{T} = \tilde{H} \), then \( x_k = y_k, k = \{1, 2, 3, 4, 6, 7, 8\} \). Therefore

\[
\frac{\sum_{k=1}^{8} x_k \times y_k}{(\sum_{k=1}^{8} x_k^2)^{\frac{1}{2}} \times (\sum_{k=1}^{8} y_k^2)^{\frac{1}{2}}} = \frac{\sum_{k=1}^{8} x_k \times x_k}{(\sum_{k=1}^{8} x_k^2)^{\frac{1}{2}} \times (\sum_{k=1}^{8} x_k^2)^{\frac{1}{2}}} = 1.
\]
Thus, 

\[ \tilde{T} = \tilde{H} \Rightarrow \cos(\tilde{T}, \tilde{H}) = 1. \]

(ii) The \( \iff \) is proved as follows: Since

\[
\cos(\tilde{T}, \tilde{H}) = 1 \Rightarrow \frac{\sum_{k=1}^{8} (x_k y_k)}{(\sum_{k=1}^{8} x_k^2)^{\frac{1}{2}} \times (\sum_{k=1}^{8} y_k^2)^{\frac{1}{2}}} = 1,
\]

\[
\sum_{k=1}^{8} (x_k y_k) = \left( \sum_{k=1}^{8} x_k^2 \right)^{\frac{1}{2}} \times \left( \sum_{k=1}^{8} y_k^2 \right)^{\frac{1}{2}}
\]

\[
(\sum_{k=1}^{8} x_k y_k)^2 = (\sum_{k=1}^{8} x_k^2) \times (\sum_{k=1}^{8} y_k^2),
\]

we have that:

\[
(\sum_{k=1}^{8} x_k^2) \times (\sum_{k=1}^{8} y_k^2)
\]

\[
= \sum_{k=1}^{8} x_k^2 \times y_k^2 + \sum_{k=1}^{8} x_k^2 \times (\sum_{q=1,q \neq k}^{8} y_q^2)
\]

\[
= \sum_{k=1}^{8} x_k^2 \times y_k^2 + \sum_{k=1}^{8} \sum_{q=1,q \neq k}^{8} x_k^2 \times y_q^2,
\]

and we have that:

\[
(\sum_{k=1}^{8} x_k y_k)^2 = \sum_{k=1}^{8} x_k^2 \times y_k^2 + \sum_{k=1}^{8} \sum_{q=1,q \neq k}^{8} x_k^2 \times y_q^2,
\]

so (a)−(b)= 0 \( \Rightarrow (x_k \times y_q) \times [(x_k \times y_q) - (x_q \times y_k)] = 0, \)

we have \( x_k \neq 0 \) and \( y_q \neq 0. \) Thus (a)−(b)= 0 \( \Rightarrow [(x_k \times y_q) - (x_q \times y_k)] = 0. \) \( \Rightarrow \exists \lambda \forall \lambda \geq 0, \) \( \frac{x_q}{y_q} = \frac{x_q}{y_q} = \lambda. \) particularly \( \lambda = 1 \Rightarrow x_k = y_k. \)

Therefore, we have that \( x_k = y_k, [k = 1, 2, 3, 4, 5, 6, 7, 8], \)
\( \tilde{T} = \tilde{H}. \)

3.4.2 MCESTA is a symmetric relation

\[ \cos(\tilde{T}, \tilde{H}) = \cos(\tilde{H}, \tilde{T}). \]

This follows since \( \sum_{k=1}^{8} x_k \times y_k = \sum_{k=1}^{8} y_k \times x_k \) and \( \sum_{k=1}^{8} x_k^2 \cdot \sum_{k=1}^{8} y_k^2 \)

\[ \Rightarrow (\sum_{k=1}^{8} x_k \times y_k)^2 = (\sum_{k=1}^{8} x_k^2)^{\frac{1}{2}} \times (\sum_{k=1}^{8} y_k^2)^{\frac{1}{2}}. \] It simple to show that
\[ \cos(\tilde{T}, \tilde{H}) = \cos(\tilde{H}, \tilde{T}). \]

3.4.3 MCESTA is scale-free

\[ \cos(\tilde{T}_1, \tilde{H}_1) = \cos(\tilde{T}, \tilde{H}). \]

Suppose \( \tilde{T}_1 = \alpha \times \tilde{T} = (\alpha \times t_1, \alpha \times t_2, \alpha \times t_3, \alpha \times t_4, \alpha \times w_{\tilde{T}}) \) and \( \tilde{H}_1 = \alpha \times \tilde{H} = (\alpha \times h_1, \alpha \times h_2, \alpha \times h_3, \alpha \times h_4, \alpha \times w_{\tilde{H}}) \)

, where \( \alpha \geq 0. \) We have
\[ \cos(\tilde{T}_1, \tilde{H}_1) = \frac{\sum_{k=1}^{8} (x_k \times y_k) \times (\alpha \times x_k) \times (\alpha \times y_k)}{\left(\sum_{k=1}^{8} (x_k \times y_k)^2\right)^{\frac{1}{2}} \times \left(\sum_{k=1}^{8} (\alpha \times x_k)^2 \times (\alpha \times y_k)^2\right)^{\frac{1}{2}}} = \cos(\tilde{T}, \tilde{H}). \]

4. A Comparative Study

4.1 Implementation

Our approach to similarity and the existing methods \( S_C \) (Eq. [1], \( S_{HC} \) (Eq. [2]), \( S_L \) (Eq. [3]), \( S_{CC} \) (Eq. [4]), \( S_{WC} \) (Eq. [5]), \( S_X \) (Eq. [6]), \( S_{S,J} \) (Eq. [7]) are implemented employing the Python 3.6 with libraries data science (NumPy, SciPy). All calculations were done with a quad-core processor 3.60 GHz and a 16 GB memory for the JVM. The sets of GTFNs are implemented as associated lists of arrays (Array List).

4.2 Examples and Comparisons

To validate and compare our contributions with the existing methods, we made the calculations on the sets already used by Hwang and Yang [10] in 2014. There are 21 sets (Set1 to Set21) of fuzzy numbers, shown in Figure[10] to Figure[30] respectively. The estimated times taken by our approach and by the existing
method are given in Table 2. We can demonstrate and analyze the weaknesses and limitations of each of the existing similarity measures from the table.

We can analyze Table 2 in terms of three types: incorrect results, scale-dependent results, and direction.

4.2.1 Incorrect results

- We have in Set1, \( S_{HC}(\tilde{A}, \tilde{A}_1) = 1 \): we can judge the result as incorrect.
Table 2. Comparison

| Method / Set | $S_L$ | $S_{HC}$ | $S_C$ | $S_{CC}$ | $S_{WC}$ | $S_X$ | $S_{SJ}$ | Our approach |
|--------------|-------|----------|-------|----------|----------|-------|----------|--------------|
| 1            | 0.9167 | 1.0      | 0.975 | 0.8357   | 0.95     | 0.9627 | 0.8356   | 0.9877       |
| 2            | 1.0    | 1.0      | 1.0   | 1.0      | 1.0      | 1.0   | 1.0      | 1.0          |
| 3            | 0.5    | 0.7692   | 0.7   | 0.42     | 0.682    | 0.7136 | 0.5997   | 0.8475       |
| 4            | 0.5    | 0.7692   | 0.7   | 0.49     | 0.7      | 0.7158 | 0.7      | 0.8551       |
| 5            | 1.0    | 1.0      | 1.0   | 0.8      | 0.8248   | 0.9652 | 0.8      | 0.9797       |
| 6            | #      | 1.0      | 1.0   | 1.0      | 1.0      | 1.0   | 1.0      | 1.0          |
| 7            | 0.0    | 0.9091   | 0.9   | 0.9      | 0.9053   | 0.9   | 0.9725   |              |
| 8            | 0.5    | 0.9091   | 0.9   | 0.54     | 0.8411   | 0.8631 | 0.5991   | 0.9482       |
| 9            | 0.6667 | 0.9091   | 0.9   | 0.81     | 0.9      | 0.9053 | 0.9      | 0.9756       |
| 10           | 0.8333 | 1.0      | 0.9   | 0.9      | 0.7833   | 0.95  | 0.8974   | 0.9311       |
| 11           | 0.75   | 1.0      | 0.9   | 0.72     | 0.8003   | 0.9127 | 0.72     | 0.9572       |
| 12           | 0.8    | 0.9375   | 0.9   | 0.78     | 0.8309   | 0.8904 | 0.8959   | 0.9068       |
| 13           | 0.75   | 0.9091   | 0.9   | 0.81     | 0.9      | 0.9053 | 0.9      | 0.979        |
| 14           | 1.0    | 1.0      | 1.0   | 0.7      | 0.7209   | 0.9553 | 0.7      | 0.9484       |
| 15           | 0.75   | 1.0      | 0.95  | 0.9048   | 0.6215   | 0.9675 | 0.9042   | 0.9187       |
| 16           | 0.4    | 0.7692   | 0.7   | 0.49     | 0.6222   | 0.7158 | 0.6971   | 0.8125       |
| 17           | 0.25   | 0.7692   | 0.7   | 0.49     | 0.7      | 0.7158 | 0.7      | 0.828        |
| 18           | 0.5    | 0.7692   | 0.7   | 0.49     | 0.7      | 0.7158 | 0.7      | 0.8551       |
| 19           | 0.5    | 0.7692   | 0.7   | 0.49     | 0.7      | 0.7158 | 0.7      | 0.7636       |
| 20           | 0.5    | 0.7692   | 0.7   | 0.49     | 0.7      | 0.7158 | 0.7      | 0.8551       |
| 21           | 0.81   | 0.8696   | 0.85  | 0.7225   | 0.85     | 0.8579 | 0.85     | 0.8551       |

Bold text represents incorrect results and italicized text, scale-dependent results.

Figure 19. Set$_{10}$.

Figure 20. Set$_{11}$.

Figure 21. Set$_{4}$.

Figure 22. Set$_{13}$.
• For Set_3 and Set_4, we have that the similarity of Set_4 should be more similar than that of Set_3, but the similarity methods $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$ and $S_L(\tilde{A}, \tilde{A}_1)$ produce the same result: we can judge the result as incorrect.

• We have in Set_5, $\tilde{A} \neq \tilde{A}_1$. The similarity produced by $S_C(\tilde{A}, \tilde{A}_1)$, $S_C(\tilde{A}, \tilde{A}_1)$ and $S_L(\tilde{A}, \tilde{A}_1)$ is 1: we can say the result is incorrect.

• We have in Set_6 for $S_L(\tilde{A}, \tilde{A}_1)$, it can’t evaluate the similarity degree. we can judge the result as incorrect.

• For Set_7, the similarity produced by $S_L(\tilde{A}, \tilde{A}_1)$ is 0: the result is incorrect.

• For Set_8 and Set_9, we have that the similarity of Set_4 should be different from the similarity of those with Set_3, but the similarity methods $S_C(\tilde{A}, \tilde{A}_1)$ and $S_C(\tilde{A}, \tilde{A}_1)$ produce the same result: we can say the result is incorrect.
• For Set$_{10}$ and Set$_{19}$, we have that the similarity of Set$_{10}$ should be more similar than that of Set$_{19}$, however, Table 2 proves that the similarity produced by the methods $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$, $S_L(\tilde{A}, \tilde{A}_1)$, $S_{CC}(\tilde{A}, \tilde{A}_1)$ and $S_X(\tilde{A}, \tilde{A}_1)$ are incorrect results.

• We have in Set$_{11}$, $\tilde{A} \neq \tilde{A}_1$. The similarity produced by $S_{HC}(\tilde{A}, \tilde{A}_1)$ is 1: we can judge the result as incorrect.

• We have in Set$_{14}$, $\tilde{A} \neq \tilde{A}_1$. The similarity produced by $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$ and $S_L(\tilde{A}, \tilde{A}_1)$ is 1: the result is incorrect.

• For Set$_{14}$ and Set$_{15}$, we have that the similarity of Set$_{14}$ should be more similar than that of Set$_{15}$. However, Table 2 proves that the similarity produced by the methods $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$, $S_L(\tilde{A}, \tilde{A}_1)$, $S_{CC}(\tilde{A}, \tilde{A}_1)$, $S_X(\tilde{A}, \tilde{A}_1)$ and $S_{JC}(\tilde{A}, \tilde{A}_1)$ are incorrect results.

• It can be seen for the two different Set$_{16}$ and Set$_{17}$ that the methods $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$, $S_{CC}(\tilde{A}, \tilde{A}_1)$ and $S_X(\tilde{A}, \tilde{A}_1)$ produce incorrect results.

• For Set$_{18}$ and Set$_{19}$, we have that the similarity of Set$_{18}$ should be more similar than that of Set$_{19}$, but the similarity methods $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$, $S_L(\tilde{A}, \tilde{A}_1)$, $S_{CC}(\tilde{A}, \tilde{A}_1)$, $S_{WC}(\tilde{A}, \tilde{A}_1)$, $S_X(\tilde{A}, \tilde{A}_1)$ and $S_{JC}(\tilde{A}, \tilde{A}_1)$ produce the same result: we can say the result is incorrect.

4.2.2 Scale-dependent results

• For Set$_{20}$ and Set$_{21}$, we have that Set$_{20}$ and Set$_{21}$ are in double-scale relation. Usually, the best similarity methods must verify property 3 (scale-free), but the similarity methods $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$, $S_L(\tilde{A}, \tilde{A}_1)$, $S_{CC}(\tilde{A}, \tilde{A}_1)$, $S_{WC}(\tilde{A}, \tilde{A}_1)$, $S_X(\tilde{A}, \tilde{A}_1)$ and $S_{JC}(\tilde{A}, \tilde{A}_1)$ are scale-dependent (are not scale-free), while our approach is scale-free.

4.2.3 Direction(Cosine) passing to large-scale

• The direction similarity is very important. This measurement is used in the big data framework when the data tends towards the infinite. But the similarity methods $S_C(\tilde{A}, \tilde{A}_1)$, $S_{HC}(\tilde{A}, \tilde{A}_1)$, $S_L(\tilde{A}, \tilde{A}_1)$, $S_{CC}(\tilde{A}, \tilde{A}_1)$, $S_{WC}(\tilde{A}, \tilde{A}_1)$, $S_X(\tilde{A}, \tilde{A}_1)$ and $S_{JC}(\tilde{A}, \tilde{A}_1)$ do not use the direction, while our approach uses this measurement with a weight $C = 0.508$, see Figure 3.

After the analysis of Table 2 the three properties and the direction based on the cosine coefficient help this similarity approach to be the best choice for the large scale.

5. Conclusion

A novel mathematical model MCESTA for GTFNs has been is presented (as well as the existing methods). This model is based on using weights associated with each one of several different similarity measures. We have been able to infer the importance weights by a Mamdani-type FIS [17]. This model uses the cosine coefficient and Jaccard Index. Three properties of the model are proved, one property is advantageous for being used with large scale datasets. A comparative study has also been presented to explain how onovel approach can overcome the limitations and weaknesses of the existing methods. This approach will help us develop an intelligent filtering of the pub-sub system [20, 21].

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

[1] L. A. Zadeh, “Fuzzy sets,” Information and Control, vol. 8, no. 3, pp. 338-353, 1965. [https://doi.org/10.1016/S0019-9958(65)90241-X]
[2] S. M. Chen, “New methods for subjective mental workload assessment and fuzzy risk analysis,” Cybernetics and Systems, vol. 27, no. 5, pp. 449-472, 1996. [https://doi.org/10.1080/019697296126417]
[3] C. H. Hsieh and S. H. Chen, “Similarity of generalized fuzzy numbers with graded mean integration representation,” in Proceedings of the 8th International Fuzzy Systems Association World Congress, 1999, Taipei, Taiwan, pp. 551-555.
[4] H. S. Lee, “An optimal aggregation method for fuzzy opinions of group decision,” in Proceedings of 1999 IEEE International Conference on Systems, Man, and Cybernetics. Tokyo, Japan, 1999, pp. 314-319. [https://doi.org/10.1109/ICSMEC.1999.823219]
[5] S. J. Chen and S. M. Chen, “Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers,” IEEE
Transactions on Fuzzy Systems, vol. 11, no. 1, pp. 45-56, 2003. [https://doi.org/10.1109/TFUZZ.2002.806316]

[6] M. S. Yang, W. L. Hung, and S. J. ChangChien, “On a similarity measure between LR-type fuzzy numbers and its application to database acquisition,” International Journal of Intelligent Systems, vol. 20, no. 10, pp. 1001-1016, 2005. [https://doi.org/10.1002/int.20102]

[7] S. H. Wei and S. M. Chen, “A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers,” Expert Systems with Applications, vol. 36, no. 1, pp. 589-598, 2009. [https://doi.org/10.1016/j.eswa.2007.09.033]

[8] Z. Xu, S. Shang, W. Qian, and W. Shu, “A method for fuzzy risk analysis based on the new similarity of trapezoidal fuzzy numbers,” Expert Systems with Applications, vol. 37, no. 3, pp. 1920-1927, 2010. [https://doi.org/10.1016/j.eswa.2009.07.015]

[9] S. J. Chen, “Fuzzy information retrieval based on a new similarity measure of generalized fuzzy numbers,” Intelligent Automation & Soft Computing, vol. 17, no. 4, pp. 465-476, 2011. [https://doi.org/10.1080/10798587.2011.1064316]

[10] C. M. Hwang and M. S. Yang, “New similarity measures between generalized trapezoidal fuzzy numbers using the Jaccard Index,” International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 22, no. 6, pp. 831-844, 2014. [https://doi.org/10.1142/S0218488514500445]

[11] S. H. Chen, “Operations of fuzzy numbers with step form membership function using function principle,” Information Sciences, vol. 108, no. 1-4, pp. 149-155, 1998. [https://doi.org/10.1016/S0020-0255(97)10070-6]

[12] K. P. Chiao, “Ranking interval type 2 fuzzy sets using parameterized graded mean integration representation,” in Proceedings of 2016 International Conference on Machine Learning and Cybernetics, Jeju, Korea, 2016, pp. 606-611. [https://doi.org/10.1109/ICMLC.2016.7872956]

[13] Y. Gupta, A. Saini, and A. K. Saxena, “Fuzzy logic-based approach to develop hybrid similarity measure for efficient information retrieval,” Journal of Information Science, vol. 40, no. 6, pp. 846-857, 2014. [https://doi.org/10.1177/0165551514548989]

[14] R. Real and J. M. Vargas, “The probabilistic basis of Jaccards index of similarity,” Systematic Biology, vol. 45, no. 3, pp. 380-385, 1996. [https://doi.org/10.1093/sysbio/45.3.380]

[15] M. Y. Samiri, M. Najib, A. Elfazziki, M. N. Abourraja, D. Boudebeou, and A. Bouain, “APRICOIN: an adaptive approach for prioritizing high-risk containers inspections,” IEEE Access, vol. 5, pp. 18238-18249, 2017. [https://doi.org/10.1109/ACCESS.2017.2746838]

[16] J. T. Tsai, J. H. Chou, and C. F. Lin, “Designing micro-structure parameters for backlight modules by using improved adaptive neuro-fuzzy inference system,” IEEE Access, vol. 3, pp. 2626-2636, 2015. [https://doi.org/10.1109/ACCESS.2015.2508144]

[17] E. H. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” International Journal of Man-Machine Studies, vol. 7, no. 1, pp. 1-13, 1975. [https://doi.org/10.1016/S0020-7373(75)80002-2]

[18] Y. Gupta, A. Saini, and A. K. Saxena, “A new fuzzy logic based ranking function for efficient information retrieval system,” Expert Systems with Applications, vol. 42, no. 3, pp. 1223-1234, 2015. [https://doi.org/10.1016/j.eswa.2014.09.009]

[19] E. Lee, C. Choi, and P. Kim, “Intelligent handover scheme for drone using fuzzy inference systems,” IEEE Access, vol. 5, pp. 13712-13719, 2017. [https://doi.org/10.1109/ACCESS.2017.2724067]

[20] M. C. Tourad and A. Abdali, “Toward efficient ranked-key algorithm for the web notification of big data systems,” in Proceedings of the 2nd International Conference on Big Data, Cloud and Applications, Tetouan, Morocco, 2017. [https://doi.org/10.1145/3090354.3090386]

[21] M. C. Tourad, A. Abdali, and A. Outfarouin, “On a new index of publish/subscribe system in the context of big data,” in Proceedings of 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications, Agadir, Morocco, 2016, pp. 1-2. [https://doi.org/10.1109/AICCSA.2016.7945756]
Mohamedou Cheikh Tourad received the degree of specialized Engineer from the Faculty of Sciences and Technology, Marrakesh, Morocco, in 2014, where he is currently pursuing a Ph.D. degree with the Computer Science Department of the Faculty of Sciences and Technology. His research interests are in Intelligent Publish/Subscribe System, Big Data and Fuzzy logic, Data science, Machine learning, Deep learning.
E-mail: cheikhtouradmohamedou@gmail.com

Abdelmounaim Abdali received a Ph.D. in Solid Mechanics and Structures from the University of Amiens, France, in 1996. He is a professor of Computer Science at the University Cadi Ayyad, Faculty of Sciences and Technology, Marrakech, Morocco, and a member of the Laboratory of Applied Mathematics and Computer Science (LAMAI) Marrakesh, Morocco. His research interests are: Publish/Subscribe System, software engineering, Big Data, Fuzzy logic, computer science, DTN network, business intelligence, design and implementation in data warehouse, ubiquitous systems, modeling & design, metamodel design, model transformation, model verification & validation methods, numerical simulation, smart cities, biomechanics, and bone remodeling.