Neutrosophic Refined Similarity Measure Based on Cosine Function

Said Broumi¹, Florentin Smarandache²

¹Faculty of Lettres and Humanities, Hay El Baraka Ben Msik Casablanca B.P. 7951, Hassan II Casablanca University, Morocco
E-mail: broumisaid78@gmail.com

²Department of Mathematics, University of New Mexico, 705 Gurley Avenue, Gallup, NM 87301, USA
E-mail: fsmarandache@gmail.com

Abstract: In this paper, the cosine similarity measure of neutrosophic refined (multi-) sets is proposed and its properties are studied. The concept of this cosine similarity measure of neutrosophic refined sets is the extension of improved cosine similarity measure of single valued neutrosophic. Finally, using this cosine similarity measure of neutrosophic refined set, the application of medical diagnosis is presented.

Keywords: Neutrosophic set, neutrosophic refined set, cosine similarity measure.

1. Introduction:

The neutrosophic sets (NS), proposed by F. Smarandache [7], has been studied and applied in different fields, including decision making problems [1,15], databases [21,22], medical diagnosis problems [2], topology [6], control theory [40], image processing [9,22,44] and so on. The concept of neutrosophic sets generalizes the following concepts: the classic set, fuzzy set [20], intuitionistic fuzzy set [19], and interval valued intuitionistic fuzzy set [18] and so on. The character of NSs is that the values of its membership function, non-membership function and indeterminacy function are subsets. Therefore, H. Wang et al [10] introduced an instance of neutrosophic sets known as single valued neutrosophic sets (SVNS), which were motivated from the practical point of view and that can be used in real scientific and engineering application, and provide the set theoretic operators and various properties of SVNSs. However, in many applications, due to lack of knowledge or data about the problem domains, the decision information may be provided with intervals, instead of real numbers. Thus, interval valued neutrosophic sets (IVNS), as a useful generation of NS, was introduced by H. Wang et al [11], which is characterized by a membership function, non-membership function and an indeterminacy function, whose values are intervals rather than real numbers. Also, the interval valued neutrosophic set can represent uncertain, imprecise, incomplete and inconsistent information which exist in the real world.

As an important extension of NS, SVNS and IVNS has many applications in real life [13,14,15,16,17,25,32,33,34,35,36,37,38,39]. Several similarity measures have been proposed by some researchers. Broumi and Smarandache [35] defined the Hausdorff distance between neutrosophic sets and some similarity measures based on the distance, set theoretic approach, and matching function to calculate the similarity degree between neutrosophic sets. In the same year, Broumi and Smarandache [32] also proposed the correlation coefficient between interval neutrosophic sets. Majumdar and Smanta [24] introduced several similarity measures of single valued neutrosophic sets (SVNs) based on distances, a matching function, membership grades, and then proposed an entropy measure for a SVNS. J. Ye [13] also presented the Hamming and Euclidean distances between interval neutrosophic sets (INSs) and their similarity measures and applied them to multiple attribute decision –making problems with interval neutrosophic information. J. Ye [15] further proposed the distance-based similarity measure of SVNSs and applied it to the group decision making problems with single valued neutrosophic information. In other research, J. Ye [16] proposed three vector similarity measure for SNSs, an instance of SVNS and INS, including the Jaccard, Dice, and cosine similarity.
measures for SVNS and INSs, and applied them to multicriteria decision-making problems with simplified neutrosophic information. Recently, A. Salama [4], introduced and studied the concepts of correlation and correlation coefficient of neutrosophic data in probability spaces and study some of their properties.

The cosine similarity measure, based on Bhattacharya’s distance [3] is the inner product of the two vectors divided by the product of their lengths. As the cosine similarity measure is the cosine of the angle between the vector representations of fuzzy sets, it is extended to cosine similarity measures between SVNSs by J. Ye [15, 17] and also to cosine similarity measures between INSs by Broumi and Smarandache [36].

The notion of multisets was formulated first in [31] by Yager as generalization of the concept of set theory. Several authors from time to time made a number of generalization of set theory. For example, Sebastian and Ramakrishnan [42] introduced a new notion called multifuzzy sets, which is a generalization of multiset. Since then, Sebastian and Ramakrishnan [41, 42] discussed more properties on multi fuzzy set. Later on, T. K. Shinoj and S. J. John [43] made an extension of the concept of fuzzy multisets by an intuitionistic fuzzy set, which called intuitionistic fuzzy multisets (IFMS). Since then in the study on IFMS, a lot of excellent results have been achieved by researchers [26, 27, 28, 29, 30]. An element of a multi fuzzy sets can occur more than once with possibly repeated occurrences of membership and non-membership. The repeated occurrence of membership and non-membership of an element of intuitionistic fuzzy multisets allows the repeated occurrences of membership and non-membership values, whereas an element of intuitionistic fuzzy multisets allows the repeated occurrences of membership and non-membership values. The concepts of FMS and IFMS fails to deal with indeterminacy. In 2013, Smarandache [8] extended the classical neutrosophic logic to n-valued refined neutrosophic logic, by refining each neutrosophic component T, I, F into respectively, T₁, T₂, ..., Tₙ and I₁, I₂, ..., Iₙ and F₁, F₂, ..., Fₙ. Recently, L. Deli et al. [12] introduced the concept of neutrosophic refined sets and studied some of their basic properties. The concept of neutrosophic refined set (NRS) is a generalization of fuzzy multisets and intuitionistic fuzzy multisets.

In this paper, motivated by the cosine similarity measure based on Bhattacharya’s distance and the improved cosine similarity measure of single valued neutrosophic proposed by J. Ye [17], we propose a new method called “cosine similarity measure for neutrosophic refined sets. The proposed cosine similarity measure is applied to medical diagnosis problems. The paper is structured as follows. In Section 2, we first recall the necessary background on cosine similarity measure and neutrosophic refined sets. In Section 3, we present cosine similarity measure for neutrosophic refined sets and examines their respective properties. In section 4, we present a medical diagnosis using NRS - cosine similarity measure. Finally we conclude the paper.

2. Preliminaries

This section gives a brief overview of the concepts of neutrosophic set, single valued neutrosophic set, cosine similarity measure and neutrosophic refined sets.

2.1 Neutrosophic Sets

Definition 2.1 [7]
Let U be an universe of discourse then the neutrosophic set A is an object having the form

\[ A = \{ x : T_A(x), I_A(x), F_A(x), x \in U \} \]

where the functions \( T, I, F : U \rightarrow [0, 1]^+ \) define respectively the degree of membership (or Truth), the degree of indeterminacy, and the degree of non-membership (or Falsehood) of the element \( x \in U \) to the set \( A \) with the condition.

\[ 0 \leq \sup (x) + \inf (x) + \sup F(x) \leq 3. \]  \hspace{1cm} (1)

From philosophical point of view, the neutrosophic set takes the value from real standard or non-standard subsets of \([0, 1]^+\). So instead of \([0, 1]^+\) we need to take the interval \([0, 1]\) for technical applications, because \([0, 1]\) will be difficult to apply in the real applications such as in scientific and engineering problems.

For two NS, \( A_{NS} = \{ < x, T_A(x), I_A(x), F_A(x) > | x \in X \} \)
And \( B_{NS} = \{ < x, T_B(x), I_B(x), F_B(x) > | x \in X \} \) the two relations are defined as follows:

\[ (1) A_{NS} \subseteq B_{NS} \text{ if and only if } (x) \leq b(x), \quad I(x) \geq I_B(x), \quad F(x) \geq F_B(x) \]

\[ (2) A_{NS} = B_{NS} \text{ if and only if, } (x) = T_B(x), \quad I(x) = I_B(x), \quad F(x) = F_B(x) \]

2.2 Single Valued Neutrosophic Sets

Definition 2.2 [10]
Let \( X \) be a space of points (objects) with generic elements in \( X \) denoted by \( x \). An SVNS A in \( X \) is characterized by a truth-membership function \( T_A(x) \), an indeterminacy-membership function \( I_A(x) \), and a falsity-membership function \( F_A(x) \), for each point \( x \) in \( X \), \( T_A(x), \quad I(x), \quad F_A(x) \in \{0, 1\} \).

When \( X \) is continuous, an SVNS A can be written as

\[ A = \int_X < T_A(x), I_A(x), F_A(x) >, x \in X. \]

(2)

When \( X \) is discrete, an SVNS A can be written as
For two SVNS, $A_{SVNS} = \{<x, T_A(x), I_A(x), F_A(x)> | x \in X\}$ and $B_{SVNS} = \{<x, T_B(x), I_B(x), F_B(x)> | x \in X\}$ the two relations are defined as follows:

1. $A_{SVNS} \subseteq B_{SVNS}$ if and only if $T_A(x) \leq T_B(x), I_A(x) \leq I_B(x), F_A(x) \leq F_B(x)$

2. $A_{SVNS} = B_{SVNS}$ if and only if, $(x) = T_B(x), I(x) = I_B(x), F_A(x) = F_B(x)$ for any $x \in X$.

### 2.3 Cosine Similarity

**Definition 2.3** [5]

Cosine similarity is a fundamental angle-based measure of similarity between two vectors of $n$ dimensions using the cosine of the angle between them. It measures the similarity between two vectors based only on the direction, ignoring the impact of the distance between them. Given two vectors of attributes, $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$, the cosine similarity, $\cos \theta$, is represented using a dot product and magnitude as

$$\cos \theta = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} y_j^2}}$$

In vector space, a cosine similarity measure based on Bhattacharya’s distance [3] between two fuzzy set $\mu_A(x_i)$ and $\mu_B(x_i)$ defined as follows:

$$C_F(A, B) = \frac{\sum_{i=1}^{n} \mu_A(x_i) \mu_B(x_i)}{\sqrt{\sum_{i=1}^{n} \mu_A(x_i)^2 \sum_{i=1}^{n} \mu_B(x_i)^2}}$$

The cosine of the angle between the vectors is within the values between 0 and 1.

In 3-D vector space, J. Ye [15] defines cosine similarity measure between SVNS as follows:

$$C_{SVNS}(A, B) = \frac{\sum_{i=1}^{n} T_A(x_i) T_B(x_i) + I_A(x_i) I_B(x_i) + F_A(x_i) F_B(x_i)}{\sqrt{\sum_{i=1}^{n} T_A(x_i)^2 + I_A(x_i)^2 + F_A(x_i)^2} \sqrt{\sum_{i=1}^{n} T_B(x_i)^2 + I_B(x_i)^2 + F_B(x_i)^2}}$$

### 2.4. Neutrosophic Refined Sets.

**Definition 2.4** [12]

Let $A$ and $B$ be two neutrosophic refined sets.

$$A = \{<x, T_A(x), I_A(x), F_A(x)> | x \in X\}$$

where $T_A(x), I_A(x), F_A(x): E \to [0, 1]$, and $F^1(x), F_A^2(x), \ldots, F_A^n(x): E \to [0, 1]$ such that $0 \leq \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \leq 3$ for $i = 1, 2, \ldots, n$ for any $x \in X$.

$.T_A(x), T_A^2(x), \ldots, T_A^n(x), (I_A^1(x), I_A^2(x), \ldots, I_A^n(x)), (F_A^1(x), F_A^2(x), \ldots, F_A^n(x))$ is the truth-membership sequence, indeterminacy-membership sequence and falsity-membership sequence of the element $x$, respectively. Also, $P$ is called the dimension of neutrosophic refined sets (NRS) $A$.

### 3. Cosine similarity measure for Neutrosophic Refined Sets.

Based on the improved cosine similarity measure of single valued neutrosophic sets proposed by J. Ye [17] which consists of membership, indeterminacy and non-membership functions defined as follow:

$$C_{SVNS}(A, B) = \frac{1}{n} \sum_{i=1}^{n} \cos \left[ \pi \left( \frac{|T_A(x_i) - T_B(x_i)| + |I_A(x_i) - I_B(x_i)| + |F_A(x_i) - F_B(x_i)|}{6} \right) \right]$$

And the cosine similarity measure of neutrosophic refined sets consisting of the multiple membership, indeterminacy, and non-membership function is

$$C_{NRS}(A, B) = \frac{1}{n} \sum_{i=1}^{n} \cos \left[ \pi \left( \frac{|I_A(x_i) - I_B(x_i)| + |F_A(x_i) - F_B(x_i)|}{6} \right) \right]$$

**Proposition 3.1.** The defined cosine similarity measure $C_{NRS}(A, B)$ between NRS $A$ and $B$ satisfies the following properties

1. $0 \leq C_{NRS}(A, B) \leq 1$
2. $C_{NRS}(A, B) = 1$ if and only if $A = B$
3. $C_{NRS}(A, B) = C_{NRS}(B, A)$
4. If $C$ is a NRS in $X$ and $A \subseteq B \subseteq C$, then $C_{NRS}(A, C) \leq C_{NRS}(A, B)$ and $C_{NRS}(A, C) \leq C_{NRS}(B, C)$

**Proof:**

(1) As the membership, indeterminacy and non-membership functions of the NRSs and the value of the cosine function are within $[0, 1]$, the similarity measure based on cosine function also is within $[0, 1]$. Hence $0 \leq C_{NRS}(A, B) \leq 1$. (2)
For any two NRSs A and B, if A=B, this implies $T_A^i(x_i) = T_B^j(x_j)$, $I_A^i(x_i) = I_B^j(x_j)$, and $F_A^i(x_i) = F_B^j(x_j)$ for $i = 1, 2, \ldots, n$ and $j=1,2,\ldots,n$ and $x_i \in X$. Hence $|T_A^i(x_i) - T_B^j(x_j)| = 0$, $|I_A^i(x_i) - I_B^j(x_j)| = 0$, and $|F_A^i(x_i) - F_B^j(x_j)| = 0$. Thus $c_{NRS}(A,B)=1$.

If $c_{NRS}(A,B)=1$ this refers that $|T_A^i(x_i) - T_B^j(x_j)| = 0$, $|I_A^i(x_i) - I_B^j(x_j)| = 0$, and $|F_A^i(x_i) - F_B^j(x_j)| = 0$ since $\cos(0)=1$. Then these equalities indicate $T_A^i(x_i) = T_B^j(x_j)$, $I_A^i(x_i) = I_B^j(x_j)$, $F_A^i(x_i) = F_B^j(x_i)$ for all $i,j$ values and $x_i \in X$. Hence $A=B$.

**Proof** is straightforward.

If $A \subseteq B \subseteq C$, then there are $T_A^i(x_i) \leq T_B^j(x_j) \leq T_C^k(x_i)$, $I_A^i(x_i) \geq I_B^j(x_j) \geq I_C^k(x_i)$, and $F_A^i(x_i) \geq F_B^j(x_i) \geq F_C^k(x_i)$ for all $i,j$ values and $x_i \in X$. Then we have the following inequalities:

$$|T_A^i(x_i) - T_B^j(x_j)| \leq |T_B^j(x_j) - T_C^k(x_i)|, |T_A^i(x_i) - T_C^k(x_i)|,$$

$$|I_A^i(x_i) - I_B^j(x_j)| \leq |I_B^j(x_j) - I_C^k(x_i)|, |I_A^i(x_i) - I_C^k(x_i)|,$$

$$|F_A^i(x_i) - F_B^j(x_j)| \leq |F_B^j(x_j) - F_C^k(x_i)|, |F_A^i(x_i) - F_C^k(x_i)|.$$

Hence, $c_{NRS}(A,C) \leq c_{NRS}(A,B)$ and $c_{NRS}(A,C) \leq c_{NRS}(B,C)$ for $k=1,2$, since the cosine function is a decreasing function within the interval $[0, \frac{\pi}{2}]$.

**4 Application**

In this section, we give some applications of NRS in medical diagnosis problems using the cosine similarity measure. Some of it is quoted from [29,30,41].

From now on, we use $A = \{\langle x, (T_A^1(x), I_A^1(x), F_A^1(x)), (T_A^2(x), I_A^2(x), F_A^2(x)), \ldots, (T_A^n(x), I_A^n(x), F_A^n(x)) \rangle : x \in X \}$.

**Table 1:** Q (the relation between patient and symptoms)

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Cough</th>
<th>Throat pain</th>
<th>Headache</th>
<th>Body pain</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>(0.4,0.3,0.4,0.5,0.4,0.4)</td>
<td>(0.3,0.4,0.6,0.4,0.1,0.3)</td>
<td>(0.2,0.5,0.5,0.1,0.6,0.3)</td>
<td>(0.5,0.3,0.4,0.5,0.3,0.4)</td>
</tr>
<tr>
<td>P2</td>
<td>(0.6,0.3,0.5,0.6,0.4,0.7)</td>
<td>(0.5,0.5,0.2,0.4,0.2,0.2)</td>
<td>(0.4,0.4,0.5,0.1,0.4,0.5)</td>
<td>(0.6,0.3,0.3,0.1,0.4,0.5)</td>
</tr>
<tr>
<td>P3</td>
<td>(0.8,0.3,0.5,0.4,0.5,0.6)</td>
<td>(0.7,0.5,0.5,0.4,0.1,0.6)</td>
<td>(0.6,0.4,0.4,0.1,0.4,0.5)</td>
<td>(0.6,0.2,0.5,0.2,0.2,0.6)</td>
</tr>
</tbody>
</table>

Let the samples be taken at three different timings in a day (in 08:00, 16:00, 24:00).

**4.1. Medical Diagnosis using NRS–cosine similarity measure**

In what follows, let us consider an illustrative example adopted from Rajarajeswari and Uma [29] with minor changes and typically considered in [30,43]. Obviously, the application is an extension of intuitionistic fuzzy multi sets [29].

"As Medical diagnosis contains lots of uncertainties and increased volume of information available to physicians from new medical technologies, the process of classifying different sets of symptoms under a single name of disease becomes difficult. In some practical situations, there is the possibility of each element having different truth membership, indeterminate and false membership functions. The proposed similarity measure among the patients Vs symptoms and symptoms Vs diseases gives the proper medical diagnosis. The unique feature of this proposed method is that it considers multi truth membership, indeterminate and false membership. By taking one time inspection, there may be error in diagnosis. Hence, this multi time inspection, by taking the samples of the same patient at different times gives best diagnosis" [29].

Now, an example of a medical diagnosis will be presented.

**Example:** Let $P=\{P_1, P_2, P_3\}$ be a set of patients, $D=\{\text{Viral Fever, Tuberculosis, Typhoid, Throat disease}\}$ be a set of diseases and $S=\{\text{Temperature, cough, throat pain, headache, body pain}\}$ be a set of symptoms. Our solution is to examine the patient at different time intervals (three times a day), which in turn give arise to different truth membership, indeterminate and false membership function for each patient.
Remark: At three different timings in a day (in 08:00, 16:00, 24:00)

P₁ upon the Temperature may have the disease 1 with chance (0.4, 0.3, 0.4) at 08:00
P₁ upon the Temperature may have the disease 2 with chance (0.3, 0.4, 0.6) at 16:00
P₁ upon the Temperature may have the disease 3 with chance (0.2, 0.5, 0.5) at 24:00

Table II: R (the relation among Symptoms and Diseases)

<table>
<thead>
<tr>
<th></th>
<th>Viral Fever</th>
<th>Tuberculosis</th>
<th>Typhoid</th>
<th>Throat disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp</td>
<td>(0.2,0.5,0.6)</td>
<td>(0.4,0.6,0.5)</td>
<td>(0.6,0.4,0.5)</td>
<td>(0.3,0.7,0.8)</td>
</tr>
<tr>
<td>Cough</td>
<td>(0.6,0.4,0.6)</td>
<td>(0.8,0.2,0.3)</td>
<td>(0.3,0.2,0.6)</td>
<td>(0.2,0.4,0.1)</td>
</tr>
<tr>
<td>Throat Pain</td>
<td>(0.5,0.2,0.3)</td>
<td>(0.4,0.5,0.5)</td>
<td>(0.4,0.5,0.5)</td>
<td>(0.2,0.6,0.2)</td>
</tr>
<tr>
<td>Headache</td>
<td>(0.6,0.8,0.2)</td>
<td>(0.2,0.3,0.6)</td>
<td>(0.1,0.6,0.3)</td>
<td>(0.2,0.5,0.5)</td>
</tr>
<tr>
<td>Body Pain</td>
<td>(0.7,0.4,0.4)</td>
<td>(0.2,0.3,0.4)</td>
<td>(0.2,0.3,0.4)</td>
<td>(0.2,0.2,0.3)</td>
</tr>
</tbody>
</table>

Table III: The Correlation Measure between NRS Q and R

<table>
<thead>
<tr>
<th></th>
<th>Cosine similarity measure</th>
<th>Viral Fever</th>
<th>Tuberculosis</th>
<th>Typhoid</th>
<th>Throat disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₁</td>
<td>0.9793</td>
<td>0.9915</td>
<td><strong>0.9896</strong></td>
<td>0.9794</td>
<td></td>
</tr>
<tr>
<td>P₂</td>
<td>0.9831</td>
<td>0.9900</td>
<td><strong>0.9870</strong></td>
<td>0.9723</td>
<td></td>
</tr>
<tr>
<td>P₃</td>
<td>0.9811</td>
<td>0.9931</td>
<td><strong>0.9917</strong></td>
<td>0.9822</td>
<td></td>
</tr>
</tbody>
</table>

The highest correlation measure from the Table III gives the proper medical diagnosis. Therefore, patient P₁, P₂ and P₃ suffers from Tuberculosis

5. Conclusion

In this paper, we have extended the improved cosine similarity of single valued neutrosophic set proposed by J.Ye [17] to the case of neutrosophic refined sets and proved some of their basic properties. We have present an application of cosine similarity measure of neutrosophic refined sets in medical diagnosis problems. In the future work, we will extend this cosine similarity measure to the case of interval neutrosophic refined sets.

Acknowledgment

The authors are very grateful to the anonymous referees for their insightful and constructive comments and suggestions, which have been very helpful in improving the paper.

References


Said Broumi, Florentin Smarandache, Neutrosophic Refined Similarity measure based on cosine function


[17] J. Ye, Improved cosine similarity measure of simplified neutrosophic sets for medicine diagnoses, artificial Intelligence in Medicine, 2014 (submitted)


Received: August 29, 2014. Accepted: October 10, 2014.

