Pathological diagnosis and clustering of single-valued neutrosophic sets based on the new similarity measure

Han-Liang Huang*
School of Mathematics and Statistics
Minnan Normal University
Zhangzhou 363000
P.R. China
hlhuang2008@163.com
hl_huang1980.student@sina.com

Xuehong Cai
Pathology department
Zhangzhou Hospital of Traditional Chinese Medicine
Zhangzhou 363000
P.R. China
3339855861@qq.com

Jiongmei Mo
School of Mathematics and Statistics
Minnan Normal University
Zhangzhou 363000
P.R. China
mojiongmei123@126.com

Abstract. Similarity measure is an effective tool for decision-making and clustering analysis. Most of the existing similarity measures only to consider the similarity between the corresponding positions of the data. On this basis, we not only consider the relationship between the corresponding positions of the data, but also consider the influence of the overall difference between the data onto the similarity in single-valued neutrosophic environment. Furthermore, similarity and distance measures with weights are given and their properties are investigated. Three examples, which include pathologic diagnosis, clustering and decision making are given to demonstrate the effectiveness of methods. It can be seen that compared with the existing methods, our methods have the characteristics of high accuracy and wide application range.

Keywords: single-valued neutrosophic set, similarity measure, distance measure, pathologic diagnosis, clustering, decision making.

1. Introduction

To handle uncertain information, Zadeh [1] presented the fuzzy set (FS), which is composed by membership function. For the voting system, there will be those who agree, disagree or abstain. In this case, it is impossible to express infor-

*. Corresponding author

mation on FS. Therefore, intuitionistic fuzzy set (IFS) which contained membership function and non-membership function is proposed by Atanassov [2]. Then, Smarandache [3] developed neutrosophic set (NS) from the philosophical standpoint, which is used to deal with indeterminate and inconsistent information. For example, in Schrödinger's Cat theory, quantum states are basically in multiple positions at the same time. They are expressed by the NS, which can represent a phenomenon that belongs both here and there at the same time. In order to simplify calculation and application convenience, Wang et al. [4] defined the special case of NS, which is called the single-valued neutrosophic set (SVNS). With the development of fuzzy information, IFS and SVNS are widely used in various fields, such as decision making [5, 6, 7, 8, 9], clustering [10, 11, 12, 13, 14] and medical diagnosis [15, 16].

Similarity measure is widely used in fuzzy information system. Different similarity measures have been studied under FS or IFS environment [17, 18, 19, 20, 21, 22, 23]. Moreover, similarity measure and correlation coefficient of SVNS have also been widely investigated [24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37]. Huang [27] introduced a new distance measure between two SVNSs for the forward and backward differences. Ye [28] developed two distance-based similarity measure of SVNS for clustering. Using the idea of priority solution, a new ranking method was proposed and applied to multi-attribute decision making by Biswas [29]. Peng [30] proposed three methods of multi-attribute decision making by defining new similarity measure and distance measure. Ye [31] introduced an improved correlation coefficient of SVNSs for multiple attribute decision making. Combined dice with cosine similarity measure, Pramanik et al. [32] defined new vector similarity measure for SVNS. Ye [33] gave a new definition of correlation coefficient of SVNS and applied it to multi-attribute decision making. Majumdar and Samanta [34] developed the distance measure and similarity measure between two SVNSs.

Pathology, as an important subject in medical science, provides an effective bridge between basic medicine and clinical medicine, and plays an important role in the theoretical research, diagnosis, treatment and prevention of diseases ([38, 39]). The basic task of surgical pathologists in the diagnosis of living tissue examination is to confirm the disease and its nature, that is, according to the organs and materials examined, first of all, to determine whether there are pathological changes, and the specific nature and type of changes, so as to help the clinical treatment and prognosis to make a judgment. At present, it is the gold standard of medical diagnosis. Classical pathological diagnosis, that is, HE slice diagnosis is mainly simple morphology, usually pathological diagnosticians directly observe and confirm the pathological tissue under naked eyes and microscopy, so as to make diagnosis according to its pathological characteristics. The intuitive features and advantages of this diagnostic method can not be replaced by other methods. However, unlike the usual medical tests, it does not reflect the indicators with precise values, but relies on the subjective experience of the examiner and lacks objective indicators. In this way, in the process of patho-

logical diagnosis, the data are highly uncertain. In the same microscopic cell morphology, different examiners sometimes to make different diagnosis, or even completely opposite diagnosis. Take soft tissue tumors as an example, the same structure can be found in different types of other tumors, such as bundle structure can be found in neoplastic fibrous tissue proliferation, fibrosarcoma, malignant peripheral nerve sheath tumor, synovial sarcoma, palisade structure occurs in neurilemmoma, malignant peripheral nerve sheath tumor, leiomyosarcoma; Pseudochrysanthemum-like and pseudochrysanthemum-like structures may appears in neuroblastoma, schwannoma, malignant peripheral neurilemmoma, etc. Different tumors often present many similar pathological structures. Fuzzy theory can effectively deal with such uncertain information ([40, 41, 42]).

In the existing literature of fuzzy similarity measurement, most of them only consider the similarity between corresponding elements, but ignore the influence of the whole data on the similarity. In order to solve this problem, a new similarity measure is presented under the single-valued neutrosophic environment in this paper, and applied to the diagnosis of soft tissue tumor and other specific examples. The main contributions of this paper are: (1) A new similarity measure for SVNS is proposed. Unlike previous definitions, we consider the effects of the differences in individual and global factors on the similarity between two SVNSs. (2) We use the defined similarity measure in pathological diagnosis. An example shows the effectiveness of the method. (3) By comparing with the existing methods of [11, 27, 28, 29, 30, 31, 32, 33], we can see that our method has higher accuracy in clustering and has a wide range of application for decision-making.

The specific framework of the article is as follows. In Section 2, we give some basic definitions. In Section 3, the new similarity measure and distance measure between two SVNSs are given. Then their properties are studied. In Section 4, three examples are given. The example of pathological diagnosis demonstrates the effectiveness of our method. The example of clustering shows the high accuracy of our method. And the example of decision making illustrates the widespread application of our method. Finally, Section 5 summarizes the paper.

2. Preliminaries

To facilitate reading and understanding, this section first reviews the concepts of NS, SVNS, entropy measure, the normalized Hamming distance and normalized Euclidian distance between two SVNSs. Finally, the abbreviations and symbols are given.

Definition 2.1 ([3]). *Let X be a space of points (objects), with a generic element in X denoted by x. A NS 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 $F_A(x)$. The functions $T_A(x)$, $I_A(x)$, and $F_A(x)$ are real standard of non-standard subsets of $]^-0, 1^+[$. That is, $T_A x : x \to]^-0, 1^+[$, $I_A x : x \to]^-0, 1^+[$*

and $T_Ax : x \to]^-0, 1^+[$. *Thus, there is no restriction on the sum of* $T_A(x)$, $I_A(x)$ *and* $F_A(x)$, *so* $^-0 \leq \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \leq 3^+$.

Definition 2.2 ([4]). *Let* X *be a space of points (objects) with generic elements in* X *denoted by* x. *A SVNS* A *in* X *is characterized by truth-membership function* $T_A(x)$, *indeterminacy-membership function* $I_A(x)$ *and falsity-membership function* $F_A(x)$. *For each point* x *in* X, $T_A(x), I_A(x), F_A(x) \in [0, 1]$. *Therefore, a SVNS* A *can be denoted by* $A = \{< x, T_A(x), I_A(x), F_A(x) > | x \in X\}$.

The following expressions are defined by Wang et al. in [4], for SVNSs A, B.

(1) $A \subseteq B$ if and only if $T_A(x) \leq T_B(x)$, $I_A(x) \geq I_B(x)$, $F_A(x) \geq F_B(x)$, for any x in X.

(2) $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

(3) $A^c = \{< x, F_A(x), 1 - I_A(x), T_A(x) > | x \in X\}$, where A^c denotes the complement of A.

Definition 2.3 ([27]). *Let* E *be a set-to-point mapping:* $E : SVNSs \to [0, 1]$, *then* E *is an entropy measure if it satisfies the following conditions:*

$(E1)$ $E(A) = 0$ *(minimum) if and only if* A *or* A^c *is a crisp set;*

$(E2)$ $E(A) = 1$ *(maximum) if and only if* $A = A^c$, *i.e.,* $T_A(x_i) = F_A(x_i)$, $I_A(x_i) = 0.5$, *for all* $x_i \in X$;

$(E3)$ $E(A) \leq E(B)$ *if* A *is less fuzzy than* B, *i.e., if* $T_A(x_i) \leq T_B(x_i), F_B(x_i) \leq F_A(x_i)$ *for* $T_B(x_i) \leq F_B(x_i)$ *and* $I_A(x_i) = I_B(x_i) = 0.5$, *or* $T_A(x_i) \geq T_B(x_i)$, $F_B(x_i) \geq F_A(x_i)$ *for* $T_B(x_i) \geq F_B(x_i)$ *and* $I_A(x_i) = I_B(x_i) = 0.5$;

$(E4)$ $E(A) = E(A^c)$.

Definition 2.4 ([34]). *Let* A, B *be two SVNSs in* $X = \{x_1, x_2, ..., x_n\}$.

(1) *The normalized Hamming distance between* A *and* B *is defined as follows:*

$$l_N(A, B) = \frac{1}{3n} \sum_{i=1}^{n} \{|T_A(x_i) - T_B(x_i)| + |I_A(x_i) - I_B(x_i)| + |F_A(x_i) - F_B(x_i)|\}.$$

(2) *The normalized Euclidian distance between* A *and* B *is defined as follows:*

$$q_N(A, B) = \Big(\frac{1}{3n} \sum_{i=1}^{n} \{(T_A(x_i) - T_B(x_i))^2 + (I_A(x_i) - I_B(x_i))^2 + (F_A(x_i) - F_B(x_i))^2\}\Big)^{\frac{1}{2}}.$$

Example 1. Let A, B, C be three SVNSs, which denoted by

$$A = \{< x_1, 0.4, 0.6, 0.8 >, < x_2, 0.2, 0.2, 0.3 >\},$$
$$B = \{< x_1, 0.5, 0.1, 0.2 >, < x_2, 0.8, 0.1, 0.1 >\},$$
$$C = \{< x_1, 0.3, 0.0, 0.6 >, < x_2, 0.4, 0.9, 0.6 >\}.$$

Then, we can calculate that

$$l_N(A, B) = \frac{1}{6}(0.1 + 0.5 + 0.6 + 0.6 + 0.1 + 0.2) = 0.35,$$

$$l_N(A,C) = \frac{1}{6}(0.1+0.6+0.2+0.2+0.7+0.3) = 0.35,$$
$$q_N(A,B) = \sqrt{\frac{1}{6}(0.01+0.25+0.36+0.36+0.01+0.04)} = \sqrt{\frac{1.03}{6}},$$
$$q_N(A,C) = \sqrt{\frac{1}{6}(0.01+0.36+0.04+0.04+0.49+0.09)} = \sqrt{\frac{1.03}{6}}.$$

We can see that l_N and q_N cannot distinguish the distance between A and B, A and C.

Abbreviations and symbols:

Fuzzy Set $\rightarrow$ FS;

Intuitionistic Fuzzy Set $\rightarrow$ IFS;

Neutrosophic Set $\rightarrow$ NS;

Single-valued Neutrosophic Set $\rightarrow$ SVNS;

X represents a space of objects;

A, B, C represent SVNSs;

$\vartheta(\vartheta^*)$ represents the single-valued neutrosophic (weighted) similarity measure;

$d(d^*)$ represents the single-valued neutrosophic (weighted) distance measure;

$I_s(I_d)$ represents the index of similarity (distance).

3. Distance measure and similarity measure

In order to improve the shortcomings of existing distance measures and consider the global impact on individual differences between set elements, we define a new similarity measure as follows.

Definition 3.1. *For two SVNSs A and B in a universe of discourse $X = \{x_1, x_2, \ldots, x_n\}$, which are denoted by $A = \{< x_i, T_A(x_i), I_A(x_i), F_A(x_i) > |x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > |x_i \in X\}$. The single-valued neutrosophic similarity measure is defined by*

$$\vartheta(A,B) = \frac{1}{3n}\sum_{i=1}^{n}\Big[\alpha_i(1-\Delta T_i) + \beta_i(1-\Delta I_i) + \gamma_i(1-\Delta F_i)\Big] \tag{1}$$

where $\Delta F_{\max} = \max_i\Big\{\Delta F_i\Big\}$

$$\alpha_i = \frac{3-\Delta T_i - \Delta T_{\max}}{3+\Delta T_{\min} - \Delta T_{\max}}, \beta_i = \frac{3-\Delta I_i - \Delta I_{\max}}{3+\Delta I_{\min} - \Delta I_{\max}}, \gamma_i = \frac{3-\Delta F_i - \Delta F_{\max}}{3+\Delta F_{\min} - \Delta F_{\max}},$$
$$\Delta T_i = \Big|T_A(x_i) - T_B(x_i)\Big|, \Delta I_i = \Big|I_A(x_i) - I_B(x_i)\Big|, \Delta F_i = \Big|F_A(x_i) - F_B(x_i)\Big|,$$
$$\Delta T_{\min} = \min_i\Big\{\Delta T_i\Big\}, \Delta T_{\max} = \max_i\Big\{\Delta T_i\Big\},$$
$$\Delta I_{\min} = \min_i\Big\{\Delta I_i\Big\}, \Delta I_{\max} = \max_i\Big\{\Delta I_i\Big\}, \Delta F_{\min} = \min_i\Big\{\Delta F_i\Big\}.$$

Definition 3.2. *Let $X = \{x_1, x_2, \ldots, x_n\}$, A and B be two SVNSs, which are denoted by $A = \{< x_i, T_A(x_i), I_A(x_i), F_A(x_i) > |x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > |x_i \in X\}$.*

$$\vartheta^*(A,B) = \frac{1}{3}\sum_{i=1}^{n}\Big\{\omega_i\Big[\alpha_i(1-\Delta T_i) + \beta_i(1-\Delta I_i) + \gamma_i(1-\Delta F_i)\Big]\Big\}, \tag{2}$$

is called a single-valued neutrosophic weighted similarity measure, where ω_i $(i = 1, 2, \ldots, n)$ are the weight of the element $x_i (i = 1, 2, \ldots, n)$ with $\omega_i \geq 0$ and $\sum_{i=1}^{n} \omega_i = 1$, the definition of $\alpha_i, \beta_i, \gamma_i, \Delta T_i, \Delta I_i$ and ΔF_i are the same as Equation (1).

Theorem 3.1. *Let A, B, C be three SVNSs in X, the similarity measure ϑ in Equation* (1) *(or ϑ^* in Equation* (2)*) has the following conditions:*

(S1) $\vartheta(A,B) = \vartheta(B,A)$;

(S2) $0 \leq \vartheta(A,B) \leq 1$;

(S3) $\vartheta(A,B) = 1$ if and only if $A = B$;

(S4) If $A \subseteq B \subseteq C$, that is $T_A(x_i) \leq T_B(x_i) \leq T_C(x_i), I_A(x_i) \geq I_B(x_i) \geq I_C(x_i), F_A(x_i) \geq F_B(x_i) \geq F_C(x_i)$, for any x_i in X, then $\vartheta(A,C) \leq \min\{\vartheta(A,B), \vartheta(B,C)\}$.

Proof of Theorem 3.1. We only prove the case of ϑ.

(S1) Obviously.

(S2) With two SVNSs $A = \{< x_i, T_A(x_i), I_A(x_i), F_A(x_i) > |x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > |x_i \in X\}$, we have $0 \leq \Delta T_i \leq 1$, $0 \leq \Delta I_i \leq 1$, $0 \leq \Delta F_i \leq 1$, which imply that $0 < \alpha_i \leq 1$, $0 < \beta_i \leq 1$ and $0 < \gamma_i \leq 1$. Then, the inequation

$$0 \leq \alpha_i(1-\Delta T_i) + \beta_i(1-\Delta I_i) + \gamma_i(1-\Delta F_i) \leq 3$$

holds. Therefore, we can prove that $0 \leq \vartheta(A,B) \leq 1$.

(S3) "$\Rightarrow$" While $\vartheta(A,B) = 1$, we have $\alpha_i(1-\Delta T_i)+\beta_i(1-\Delta I_i)+\gamma_i(1-\Delta F_i) = 3$ for any $i = 1, 2, \ldots, n$. Due to $\alpha_i \leq 1$, $\beta_i \leq 1$, $\gamma_i \leq 1$,$1-\Delta T_i \leq 1$, $1-\Delta I_i \leq 1$ and $1-\Delta F_i \leq 1$, we can get $\Delta T_i = \Delta I_i = \Delta F_i = 0$ for all i. This implies that $A = B$.

"$\Leftarrow$" Suppose $A = B$, that is $T_A(x_i) = T_B(x_i)$, $I_A(x_i) = I_B(x_i)$ and $F_A(x_i) = F_B(x_i)$. Further, it holds that $\Delta T_i = \Delta I_i = \Delta F_i = 0$. Then $\alpha_i = \beta_i = \gamma_i = 1$ for all i. So we have $\alpha_i(1-\Delta T_i) + \beta_i(1-\Delta I_i) + \gamma_i(1-\Delta F_i) = 3$. Therefore, $\vartheta(A,B) = 1$.

(S4) Suppose $A \subseteq B \subseteq C$, that is $T_A(x_i) \leq T_B(x_i) \leq T_C(x_i)$, $I_A(x_i) \geq I_B(x_i) \geq I_C(x_i)$, $F_A(x_i) \geq F_B(x_i) \geq F_C(x_i)$. On one hand, we have the following inequations for all $i = 1, 2, \ldots, n$,

$$\begin{aligned}
&\Delta T_i^{AC} \geq \Delta T_i^{AB} \Rightarrow \Delta T_{\min}^{AC} \geq \Delta T_{\min}^{AB}, \quad \Delta T_{\max}^{AC} \geq \Delta T_{\max}^{AB},\\
&\Delta I_i^{AC} \geq \Delta I_i^{AB} \Rightarrow \Delta I_{\min}^{AC} \geq \Delta I_{\min}^{AB}, \quad \Delta I_{\max}^{AC} \geq \Delta I_{\max}^{AB},\\
&\Delta F_i^{AC} \geq \Delta F_i^{AB} \Rightarrow \Delta F_{\min}^{AC} \geq \Delta F_{\min}^{AB}, \quad \Delta F_{\max}^{AC} \geq \Delta F_{\max}^{AB}.
\end{aligned}$$

This implies that $1-\Delta T_i^{AC} \leq 1-\Delta T_i^{BC}$, $1-\Delta I_i^{AC} \leq 1-\Delta I_i^{BC}$ and $1-\Delta F_i^{AC} \leq 1-\Delta F_i^{BC}$.

On the other hand, let $F(x,y,z)=\frac{3-x-y}{3+z-y}$. Then calculate the partial derivative of F on x,y,z, respectively, we have

$$F_x=\frac{-1}{3+z-y}<0,\ F_y=\frac{-z-x}{(3+z-y)^2}\leq 0, F_z=\frac{-3+x+y}{(3+z-y)^2}<0.$$

We know that F is monotonously decreasing on x,y,z. Thus, $\alpha_i^{AC}\leq\alpha_i^{AB}$, $\beta_i^{AC}\leq\beta_i^{AB}$ and $\gamma_i^{AC}\leq\gamma_i^{AB}$.

By Equation (1), we can get $\vartheta(A,C)\leq\vartheta(A,B)$. Similarly, we can prove $\vartheta(A,C)\leq\vartheta(B,C)$. Finally, $\vartheta(A,C)\leq\min\{\vartheta(A,B),\vartheta(B,C)\}$ holds for $A\subseteq B\subseteq C$.

Theorem 3.2. *Let A be a SVNS, then $\vartheta(A,A^c)$ is an entropy measure for A.*

Proof of Theorem 3.2. Let $E(A)=\vartheta(A,A^c)$ and need to verify it satisfies the four conditions in Definition 2.3. Suppose $A=\{<x_i,T_A(x_i),I_A(x_i),F_A(x_i)>|x_i\in X\}$ and $B=\{<x_i,T_B(x_i),I_B(x_i),F_B(x_i)>|x_i\in X\}$.

(E1) "$\Rightarrow$" When $\vartheta(A,A^c)=0$, due to $\alpha_i>0$, $\beta_i>0$, $\gamma_i>0$, we have $1-\Delta T_i^{AA^c}=1-\Delta I_i^{AA^c}=1-\Delta F_i^{AA^c}=0$ for any $i=1,2,\ldots,n$. Then, $\Delta T_i^{AA^c}=\Delta I_i^{AA^c}=\Delta F_i^{AA^c}=1$ which implies that $T_A(x_i)=1,I_A(x_i)=F_A(x_i)=0$ or $T_A(x_i)=I_A(x_i)=0,F_A(x_i)=1$ or $T_A(x_i)=I_A(x_i)=1,F_A(x_i)=0$ or $T_A(x_i)=0,I_A(x_i)=F_A(x_i)=1$ must hold for all $x_i\in X$. This shows that A or A^c is a crisp set.

"$\Leftarrow$" Assume that A is a crisp set, i.e. $T_A(x_i)=1,I_A(x_i)=F_A(x_i)=0$ or $F_A(x_i)=1,T_A(x_i)=I_A(x_i)=0$ for all $x_i\in X$. No matter in which case, we have $\vartheta(A,A^c)=0$. Analogously we can prove that $\vartheta(A,A^c)=0$, when A^c is a crisp set.

(E2) By the condition (3) in Theorem 3.1, it holds that $\vartheta(A,A^c)=1\Leftrightarrow A=A^c$.

(E3) Suppose that $T_A(x_i)\leq T_B(x_i),F_B(x_i)\leq F_A(x_i)$ for $T_B(x_i)\leq F_B(x_i)$ and $I_A(x_i)=I_B(x_i)=0.5$, we have the following results,

$$\begin{aligned}
&\Delta T_i^{AA^c}\geq\Delta T_i^{BB^c}\Rightarrow\alpha_i^{AA^c}\leq\alpha_i^{BB^c},\quad 1-T_i^{AA^c}\leq 1-\Delta T_i^{BB^c},\\
&\Delta F_i^{AA^c}\geq\Delta F_i^{BB^c}\Rightarrow\gamma_i^{AA^c}\leq\gamma_i^{BB^c},\quad 1-F_i^{AA^c}\leq 1-\Delta F_i^{BB^c},\\
&\Delta I_i^{AA^c}=\Delta I_i^{BB^c}=0\Rightarrow\beta_i^{AA^c}=\beta_i^{BB^c}=1,\quad 1-\Delta I_i^{AA^c}=1-\Delta I_i^{BB^c}=1.
\end{aligned}$$

This shows that $E(A)=\vartheta(A,A^c)\leq\vartheta(B,B^c)=E(B)$. Analogously we can prove that $\vartheta(A,A^c)\leq\vartheta(B,B^c)$ when $T_A(x_i)\geq T_B(x_i),F_B(x_i)\geq F_A(x_i)$ for $T_B(x_i)\geq F_B(x_i)$ and $I_A(x_i)=I_B(x_i)=0.5$.

(E4) By the condition (1) in Theorem 3.1, we have $E(A)=\vartheta(A,A^c)=\vartheta(A^c,A)=\vartheta(A^c,A^{cc})=E(A^c)$.

In some cases, we should consider not only the similarity between two SVNSs A and B, but also the similarity between A and B^c. The index of similarity for SVNSs A and B is defined in Definition 3.3.

Definition 3.3. *Let A and B be two SVNSs, then*

$$I_s(A,B) = \frac{\vartheta(A,B)}{\vartheta(A,B^c)}$$

is called the index of similarity.

Proposition 3.1. *The index of similarity $I_s(A,B)$ for two SVNSs A and B has the following properties:*

$(IS1)$ $I_s(A,B) = 0$ means A and B are completely dissimilar;

$(IS2)$ $I_s(A,B) \to +\infty$ means A and B^c are completely dissimilar;

$(IS3)$ $I_s(A,B) = 1$, i.e. $\vartheta(A,B) = \vartheta(A,B^c)$, this means the similarity between A and B, A and B^c are the same;

$(IS4)$ $I_s(A,B) < 1$ means compare with B^c, A is less similar to B;

$(IS5)$ $I_s(A,B) > 1$ means compare with B^c, A is more similar to B.

Definition 3.4. *Let $X = \{x_1, x_2, \ldots, x_n\}$, A and B be two SVNSs, which are denoted by $A = \{< x_i, T_A(x_i), I_A(x_i), F_A(x_i) > | x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > | x_i \in X\}$. The single-valued neutrosophic distance measure are defined by*

$$d(A,B) = 1 - \vartheta(A,B), \tag{3}$$

where $\vartheta(A,B)$ is defined as Equation (1).

Remark. Using Equation 3 to calculate the distance between SVNSs A and B, A and C which are given in Example 1.

$$\begin{aligned} d(A,B) &= 1 - \frac{1}{6}\Big(\frac{23}{25}\times\frac{9}{10}+\frac{20}{26}\times\frac{5}{10}+\frac{18}{26}\times\frac{4}{10}+\frac{18}{25}\times\frac{4}{10}+\frac{24}{26}\times\frac{9}{10}+\frac{22}{26}\times\frac{8}{10}\Big) \\ &= 0.4525, \\ d(A,C) &= 1 - \frac{1}{6}\Big(\frac{27}{29}\times\frac{9}{10}+\frac{17}{29}\times\frac{4}{10}+\frac{25}{29}\times\frac{8}{10}+\frac{26}{29}\times\frac{8}{10}+\frac{16}{29}\times\frac{3}{10}+\frac{24}{29}\times\frac{7}{10}\Big) \\ &= 0.4623. \end{aligned}$$

We can see that A is closer to B.

Definition 3.5. *Let $X = \{x_1, x_2, \ldots, x_n\}$, A and B be two SVNSs, which are denoted by $A = \{< x_i,\ T_A(x_i), I_A(x_i), F_A(x_i) > | x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > | x_i \in X\}$. The single-valued neutrosophic weighted distance measure are defined by*

$$d^*(A,B) = 1 - \vartheta^*(A,B), \tag{4}$$

where $\vartheta^*(A,B)$ is defined as Equation (2).

Theorem 3.3. *Let A, B, C be three SVNSs in X, the distance measure d in Equation* (3) (*or d^* in Equation* (4)) *satisfies the following properties:*

$(D1)$ $d(A, B) = d(B, A)$;

$(D2)$ $0 \leq d(A, B) \leq 1$;

$(D3)$ $d(A, B) = 0$ *if and only if* $A = B$;

$(D4)$ *If* $A \subseteq B \subseteq C$, *that is* $T_A(x_i) \leq T_B(x_i) \leq T_C(x_i)$, $I_A(x_i) \geq I_B(x_i) \geq I_C(x_i), F_A(x_i) \geq F_B(x_i) \geq F_C(x_i)$ *for any* x_i *in* X, *then* $d(A, C) \geq \max\{d(A, B), d(B, C)\}$.

Definition 3.6. *For two SVNSs A and B,*

$$I_d(A, B) = \frac{d(A, B)}{d(A, B^c)}$$

is called the index of distance.

Proposition 3.2. *Let A and B be two SVNSs, then the index of distance $I_d(A, B)$ has the following properties:*

$(ID1)$ $I_d(A, B) = 0$ *if and only if* $A = B$;

$(ID2)$ $I_d(A, B) \to +\infty$, *i.e.* $A = B^c$, *this means A and B^c are completely different;*

$(ID3)$ $I_d(A, B) = 1$, *i.e.* $d(A, B) = d(A, B^c)$, *this means the distance between A and B, A and B^c are the same;*

$(ID4)$ $I_d(A, B) < 1$ *means compare with B^c, A is less different to B;*

$(ID5)$ $I_d(A, B) > 1$ *means compare with B^c, A is more different to B;*

$(ID6)$ *If* $I_d(A, B) = \frac{0}{0}$, *i.e.* $A = B = B^c$, *then the entropy measure of A and B reaches its maximum value.*

4. Application examples

4.1 Pathological diagnosis

There are many uncertainties of pathological diagnosis. We will give the general steps and an example of pathological diagnosis under the SVNS environment.

For the set of pathological diagnosis results $A = \{A_1, A_2, \ldots, A_n\}$, the set of patients $B = \{B_1, B_2, \ldots, B_m\}$, and the set of microscopic shape results $C = \{C_1, C_2, \ldots, C_t\}$, the diagnosis steps are as follows:

Step 1. According to the doctors' experiences, each microscopic shape result of each pathological diagnosis result is described by a single-valued neutrosophic number. Then we have a matrix $M = (M_{ki})_{t\times n} = (M_1, M_2, \ldots, M_n)$;

Step 2. For every patient $B_j (j = 1, 2, \ldots, m)$, the microscopic shape result by examiner is also described by a single-valued neutrosophic number. Then we have a matrix $N = (N_{kj})_{t\times m} = (N_1, N_2, \ldots, N_m)$;

Step 3. Use the proposed similarity measure (1) to calculate $\vartheta(M_i, N_j), (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$. Then we have the final matrix $D = D(ij)_{n\times m} = (D_1, D_2, \ldots, D_m)$;

Step 4. For every patient $B_j (j = 1, 2, \ldots, m)$, select the diagnosis result $D_j^* = \max_i D_{ij}$, $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, m)$.

Example 2. To make a proper pathological diagnosis, we have the following sets. The set of pathological diagnosis results is

$$\begin{aligned} A &= \{A_1, A_2, A_3, A_4, A_5\} \\ &= \{\text{Fibrosarcoma, Malignant fibrous histiocytoma,} \\ &\quad \text{Spindle celltype synovial sarcoma,} \\ &\quad \text{Schwannoma, Malignant peripheral nerve sheath tumor}\}. \end{aligned}$$

The set of patients is $B = \{B_1, B_2, B_3, B_4, B_5\} = \{Au, Hung, Jo, Keo, Pao\}$. The set of microscopic shape results is

$$\begin{aligned} C &= \{C_1, C_2, C_3, C_4, C_5, C_6, C_7\} \\ &= \{\text{Palisade structure, Fishbone like structure, Spoke structure,} \\ &\quad \text{Fascicular structure,} \\ &\quad \text{Multiple nuclear fission, Pleomorphic, Fish-meat like}\}. \end{aligned}$$

The data are given in Table 1 and Table 2.

Table 1: The characteristics of each microscopic shape of the pathological diagnosis

	A_1	A_2	A_3	A_4	A_5
C_1	$<0.1, 0.2, 0.8>$	$<0.1, 0.3, 0.8>$	$<0.4, 0.3, 0.5>$	$<0.9, 0.1, 0.1>$	$<0.3, 0.2, 0.8>$
C_2	$<0.8, 0.1, 0.2>$	$<0.5, 0.1, 0.4>$	$<0.7, 0.1, 0.2>$	$<0.3, 0.2, 0.6>$	$<0.4, 0.1, 0.5>$
C_3	$<0.1, 0.3, 0.7>$	$<0.9, 0.1, 0.0>$	$<0.3, 0.2, 0.6>$	$<0.1, 0.2, 0.9>$	$<0.2, 0.1, 0.7>$
C_4	$<0.9, 0.1, 0.2>$	$<0.8, 0.1, 0.3>$	$<0.5, 0.1, 0.3>$	$<0.6, 0.3, 0.3>$	$<0.8, 0.2, 0.2>$
C_5	$<0.8, 0.2, 0.3>$	$<0.7, 0.2, 0.2>$	$<0.9, 0.1, 0.2>$	$<0.3, 0.2, 0.7>$	$<0.8, 0.2, 0.2>$
C_6	$<0.3, 0.1, 0.6>$	$<0.8, 0.3, 0.1>$	$<0.2, 0.2, 0.8>$	$<0.2, 0.1, 0.8>$	$<0.7, 0.1, 0.2>$
C_7	$<0.7, 0.3, 0.2>$	$<0.1, 0.1, 0.9>$	$<0.9, 0.1, 0.1>$	$<0.1, 0.3, 0.8>$	$<0.8, 0.2, 0.1>$

Table 2: The characteristics of each microscopic shape of the considered patients

	B_1	B_2	B_3	B_4	B_5
C_1	$<0.7, 0.2, 0.2>$	$<0.3, 0.2, 0.6>$	$<0.4, 0.2, 0.6>$	$<0.1, 0.1, 0.8>$	$<0.3, 0.1, 0.7>$
C_2	$<0.1, 0.3, 0.6>$	$<0.6, 0.2, 0.3>$	$<0.5, 0.2, 0.3>$	$<0.7, 0.2, 0.2>$	$<0.5, 0.2, 0.3>$
C_3	$<0.3, 0.2, 0.6>$	$<0.3, 0.2, 0.8>$	$<0.1, 0.3, 0.7>$	$<0.3, 0.3, 0.4>$	$<0.7, 0.1, 0.2>$
C_4	$<0.8, 0.1, 0.1>$	$<0.5, 0.2, 0.4>$	$<0.2, 0.2, 0.6>$	$<0.7, 0.3, 0.1>$	$<0.5, 0.2, 0.2>$
C_5	$<0.2, 0.2, 0.7>$	$<0.9, 0.1, 0.1>$	$<0.8, 0.2, 0.1>$	$<0.8, 0.2, 0.1>$	$<0.7, 0.3, 0.2>$
C_6	$<0.4, 0.2, 0.4>$	$<0.7, 0.1, 0.5>$	$<0.4, 0.2, 0.7>$	$<0.5, 0.3, 0.3>$	$<0.9, 0.2, 0.2>$
C_7	$<0.2, 0.1, 0.7>$	$<0.4, 0.2, 0.4>$	$<0.6, 0.1, 0.3>$	$<0.8, 0.1, 0.2>$	$<0.5, 0.1, 0.7>$

Then we use the proposed similarity measure (1) and have the pathological diagnosis results of the considered patients in Table 3.

Table 3: Similarity measure of pathological diagnosis results of each patient

	B_1(Au)	B_2(Hung)	B_3(Jo)	B_4(Keo)	B_5(Pao)
A_1 (Fibrosarcoma)	0.6524	0.7747	0.8131	**0.8467**	0.6857
A_2 (Malignant fibrous histiocytoma)	0.6533	0.6855	0.6524	0.7342	**0.8400**
A_3 (Spindle celltype synovial sarcoma)	0.6786	0.8157	**0.8177**	0.8044	0.6955
A_4 (Schwannoma)	**0.8070**	0.6724	0.6626	0.5805	0.6110
A_5 (Malignant peripheral nerve sheath tumor)	0.6927	**0.8199**	0.7884	0.8336	0.7968

Finally, we have the proper pathological diagnosis as follows:

Au got schwannoma, Hung got malignant peripheral nerve sheath tumor, Jo got spindle celltype synovial sarcoma, Keo got fibrosarcoma, Pao got malignant fibrous histiocytoma.

4.2 Clustering

Under the single-valued neutrosophic data environment, we apply the proposed similarity measure to cluster by using the method of [28]. The method is described as follows:

Step 1. Using the Equations (1) (or (2)), we can calculate a similarity matrix $C = (s_{ij})_{m\times m}$, where $s_{ij} = s_{ji} = \vartheta(A_i, A_j)$ for $i, j = 1, 2, \ldots, m$.

Step 2. Repeat the following operations

$$C \to C^2 \to C^4 \to \ldots \to C^{2^k}$$

until $C^{2^k} = C^{2^{k+1}}$, where $C^2 = C \circ C = (s'_{ij})_{m\times m} = \max_k\{\min(s_{ik}, s_{kj})\}_{m\times m}(i, j = 1, 2, \ldots, m)$. Then, we have an equivalent matrix C^{2^k}.

Step 3. Calculate the α-cutting matrix $\bar{C}_\alpha = (\bar{s}^\alpha_{ij})_{m\times m}$ of $\bar{C} = C^{2^k}$ by

$$\bar{s}^\alpha_{ij} = \begin{cases} 0, & \bar{s}_{ij} < \alpha; \\ 1, & \bar{s}_{ij} \geq \alpha, \end{cases} \tag{5}$$

for $j, j = 1, 2, \ldots, m$ and α is the confidence level with $\alpha \in [0, 1]$.

Step 4. Choose different values of α, then we can classify A_i. If $\bar{s}^\alpha_{ij} = \bar{s}^\alpha_{kj}$ for all $j = 1, 2, \ldots m$, then we can put A_i and A_k to the same class.

In the following, we consider a numerical example which is given by Ye [28].

Example 3 [28] A car market is going to classify five different cars of A_m $(m = 1, 2, 3, 4, 5)$. Every car has six evaluation factors (attributes): (i) x_1, fuel consumption; (ii) x_2, coefficient of friction; (iii) x_3, price; (iv) x_4, comfortable degree; (v) x_5, design; (vi) x_6, security coefficient. The characteristics of each car under the six attributes are represented by the form of SVNSs, and then the single-valued neutrosophic data are in Table 4:

Table 4: The characteristics of each car under the attributes are expressed by SVNSs

	x_1	x_2	x_3	x_4	x_5	x_6
A_1	$<0.3,0.2,0.5>$	$<0.6,0.3,0.1>$	$<0.4,0.3,0.3>$	$<0.8,0.1,0.1>$	$<0.1,0.3,0.6>$	$<0.5,0.2,0.4>$
A_2	$<0.6,0.3,0.3>$	$<0.5,0.4,0.2>$	$<0.6,0.2,0.1>$	$<0.7,0.2,0.1>$	$<0.3,0.1,0.6>$	$<0.4,0.3,0.3>$
A_3	$<0.4,0.2,0.4>$	$<0.8,0.2,0.1>$	$<0.5,0.3,0.1>$	$<0.6,0.1,0.2>$	$<0.4,0.1,0.5>$	$<0.3,0.2,0.2>$
A_4	$<0.2,0.4,0.4>$	$<0.4,0.5,0.1>$	$<0.9,0.2,0.0>$	$<0.8,0.2,0.1>$	$<0.2,0.3,0.5>$	$<0.7,0.3,0.1>$
A_5	$<0.5,0.3,0.2>$	$<0.3,0.2,0.6>$	$<0.6,0.1,0.3>$	$<0.7,0.1,0.1>$	$<0.6,0.2,0.2>$	$<0.5,0.2,0.3>$

Firstly, we use the similarity measure formula (1) to calculate the similarity measures between each pair of SVNSs A_m $(m = 1, 2, \ldots, 5)$. For example, we use ΔT_1^{12} to express $\Delta T_1^{A_1,A_2}$, then

$$\begin{aligned}
&\Delta T_1^{12} = 0.3,\ \Delta T_2^{12} = 0.1,\ \Delta T_3^{12} = 0.2,\ \Delta T_4^{12} = 0.1,\\
&\Delta T_5^{12} = 0.2,\ \Delta T_6^{12} = 0.1, \Delta I_1^{12} = 0.1,\ \Delta I_2^{12} = 0.1,\\
&\Delta I_3^{12} = 0.1,\ \Delta I_4^{12} = 0.1,\ \Delta I_5^{12} = 0.2,\ \Delta I_6^{12} = 0.1,\\
&\Delta F_1^{12} = 0.2,\ \Delta F_2^{12} = 0.1,\ \Delta F_3^{12} = 0.2,\ \Delta F_4^{12} = 0,\\
&\Delta F_5^{12} = 0,\ \Delta F_6^{12} = 0.1, \Delta T_{\min}^{12} = 0.1,\ \Delta T_{\max}^{12} = 0.3,\\
&\Delta I_{\min}^{12} = 0.1,\ \Delta I_{\max}^{12} = 0.2,\ \Delta F_{\min}^{12} = 0,\ \Delta F_{\max}^{12} = 0.2.
\end{aligned}$$

By Definition 3.1, we have $\alpha_1^{12} = \frac{3-0.3-0.3}{3+0.1-0.3} = \frac{24}{28}$, analogously,

$$\begin{aligned}
&\alpha_2^{12} = \frac{26}{28},\ \alpha_3^{12} = \frac{25}{28},\ \alpha_4^{12} = \frac{26}{28},\ \alpha_5^{12} = \frac{25}{28},\ \alpha_6^{12} = \frac{26}{28},\\
&\beta_1^{12} = \frac{27}{29},\ \beta_2^{12} = \frac{27}{29},\ \beta_3^{12} = \frac{27}{29},\ \beta_4^{12} = \frac{27}{29},\ \beta_5^{12} = \frac{26}{29},\ \beta_6^{12} = \frac{27}{29},\\
&\gamma_1^{12} = \frac{26}{28},\ \gamma_2^{12} = \frac{27}{28},\ \gamma_3^{12} = \frac{26}{28},\ \gamma_4^{12} = 1,\ \gamma_5^{12} = 1, \gamma_6^{12} = \frac{27}{28}.
\end{aligned}$$

Furthermore, we can get $\vartheta(A_1, A_2) = \frac{1}{18}\{\frac{24}{28}\frac{7}{10}+\frac{26}{28}\frac{9}{10}+\frac{25}{28}\frac{8}{10}+\frac{26}{28}\frac{9}{10}+\frac{25}{28}\frac{8}{10}+\frac{26}{28}\frac{9}{10}+\frac{27}{29}\frac{9}{10}+\frac{27}{29}\frac{9}{10}+\frac{27}{29}\frac{9}{10}+\frac{27}{29}\frac{9}{10}+\frac{26}{29}\frac{8}{10}+\frac{27}{29}\frac{9}{10}+\frac{26}{28}\frac{8}{10}+\frac{27}{28}\frac{9}{10}+\frac{26}{28}\frac{8}{10}+1+1+\frac{27}{28}\frac{9}{10}\} = 0.8147$.

The similarity measures between each A_i and A_j $(i, j = 1, 2, 3, 4, 5, i \neq j)$ are shown as follows:

$$\begin{aligned}
&\vartheta(A_1, A_2) = 0.8147, \quad \vartheta(A_1, A_3) = 0.8464,\\
&\vartheta(A_1, A_4) = 0.8148, \quad \vartheta(A_1, A_5) = 0.7824,\\
&\vartheta(A_2, A_3) = 0.8453, \quad \vartheta(A_2, A_4) = 0.8264,\\
&\vartheta(A_2, A_5) = 0.8294, \quad \vartheta(A_3, A_4) = 0.7563,\\
&\vartheta(A_3, A_5) = 0.7680, \quad \vartheta(A_4, A_5) = 0.7190.
\end{aligned}$$

Then, we have a similarity matrix as follows:

$$C = \begin{bmatrix} 1 & 0.8147 & 0.8464 & 0.8148 & 0.7824 \\ 0.8147 & 1 & 0.8453 & 0.8264 & 0.8294 \\ 0.8464 & 0.8453 & 1 & 0.7563 & 0.7680 \\ 0.8148 & 0.8264 & 0.7563 & 1 & 0.7190 \\ 0.7824 & 0.8294 & 0.7680 & 0.7190 & 1 \end{bmatrix}.$$

Then, by Step 2 in Chapter 4.2, we have

$$C^2 = \begin{bmatrix} 1 & 0.8453 & 0.8464 & 0.8148 & 0.8147 \\ 0.8453 & 1 & 0.8453 & 0.8264 & 0.8294 \\ 0.8464 & 0.8453 & 1 & 0.8264 & 0.8294 \\ 0.8148 & 0.8264 & 0.8264 & 1 & 0.8264 \\ 0.8147 & 0.8294 & 0.8294 & 0.8264 & 1 \end{bmatrix}.$$

Due to $C^2 \nsubseteq C$, i.e., C is not an equivalent matrix, we keep calculating.

$$C^4 = \begin{bmatrix} 1 & 0.8453 & 0.8464 & 0.8264 & 0.8294 \\ 0.8453 & 1 & 0.8453 & 0.8264 & 0.8294 \\ 0.8464 & 0.8453 & 1 & 0.8264 & 0.8294 \\ 0.8264 & 0.8264 & 0.8264 & 1 & 0.8264 \\ 0.8294 & 0.8294 & 0.8294 & 0.8264 & 1 \end{bmatrix}.$$

$$C^8 = \begin{bmatrix} 1 & 0.8453 & 0.8464 & 0.8264 & 0.8294 \\ 0.8453 & 1 & 0.8453 & 0.8264 & 0.8294 \\ 0.8464 & 0.8453 & 1 & 0.8264 & 0.8294 \\ 0.8264 & 0.8264 & 0.8264 & 1 & 0.8264 \\ 0.8294 & 0.8294 & 0.8294 & 0.8264 & 1 \end{bmatrix}.$$

It is clearly that $C^4 \nsubseteq C^2$, but $C^8 = C^4$. That is, C^4 is an equivalent matrix, denoted by $\bar{C}$.

Finally, we calculate the α-cutting matrix $\bar{C}_\alpha$ by setting different values of α.

(I) Let $0 \leq \alpha \leq 0.8264$,

$$\bar{C}_\alpha = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

By Step 4 in Section 4.2, A_m $(m = 1, 2, \ldots, 5)$ can be divided into one category $\{A_1, A_2, A_3, A_4, A_5\}$.

(II) Let $0.8264 < \alpha \leq 0.8294$,

$$\bar{C}_\alpha = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

Then, the cars A_m $(m = 1, 2, \ldots, 5)$ can be divided into two categories $\{A_1, A_2, A_3, A_5\}$, $\{A_4\}$.

(III) Let $0.8294 < \alpha \leq 0.8453$,

$$\bar{C}_\alpha = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then, the cars A_m $(m = 1, 2, \ldots, 5)$ can be divided into three categories $\{A_1, A_2, A_3\}$, $\{A_4\}$, $\{A_5\}$.

(IV) Let $0.8453 < \alpha \leq 0.8464$,

$$\bar{C}_\alpha = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then the cars A_m $(m = 1, 2, \ldots, 5)$ can be divided into four categories $\{A_1, A_3\}$, $\{A_2\}$, $\{A_4\}$, $\{A_5\}$.

(V) Let $0.8464 < \alpha \leq 1$,

$$\bar{C}_\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then the cars A_m $(m = 1, 2, \ldots, 5)$ can be divided into five categories $\{A_1\}$, $\{A_2\}$, $\{A_3\}$, $\{A_4\}$, $\{A_5\}$.

Comparative analysis of existing methods. Refs. [27, 28, 11] considered the same problem as Example 2, the results are shown in Table 5. For convenience, we use the corresponding subscript m to represent A_m, $(m = 1, 2, 3, 4, 5)$.

Table 5: Clustering results of different Refs.

	Method from [27]	Method from [28]	Method from [11]	our method
1 category	$\{1, 2, 3, 4, 5\}$	$\{1, 2, 3, 4, 5\}$	$\{1, 2, 3, 4, 5\}$	$\{1, 2, 3, 4, 5\}$
2 categories	$\{1, 2, 3, 5\}, \{4\}$	$\{1, 2, 3, 4\}, \{5\}$	failed	$\{1, 2, 3, 5\}, \{4\}$
3 categories	$\{1, 2, 3\}, \{4\}, \{5\}$	$\{1, 2, 3\}, \{4\}, \{5\}$	$\{1, 2, 3\}, \{4\}, \{5\}$	$\{1, 2, 3\}, \{4\}, \{5\}$
4 categories	$\{1\}, \{2, 3\}, \{4\}, \{5\}$	$\{1\}, \{2, 3\}, \{4\}, \{5\}$	failed	$\{1, 3\}, \{2\}, \{4\}, \{5\}$
5 categories	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$

By comparison, we can see the superiority of our results as following.

(1) Ref. [11] can only classify objectives into one category, three categories and five categories, but can not classify them into two categories and four categories. However, our approach divides objectives into one to five categories. This means that our method has better accuracy in some clustering problems.

(2) Our clustering results are also different from the results of [27, 28]. In [27], most classifications are the same as our results except dividing the cars into four categories $\{A_1\}$, $\{A_2, A_3\}$, $\{A_4\}$, $\{A_5\}$. In [28], the difference only appeared in dividing the cars into two categories $\{A_1, A_2, A_3, A_4\}$, $\{A_5\}$. The reason is that the similarity measure or the distance measure in [27, 28] only consider about the difference between each corresponding data, but ignore the impact on the overall data gap between each data.

4.3 Decision making

Example 3 ([8])**.** A city is going to build a municipal library. How to determine what kind of air-conditioning system should be installed in the library is one of the problems facing the city development commissioner. The contractor provides five feasible alternatives A_i $(i = 1, 2, 3, 4, 5)$, which might be adapted to the library. Supposing that three attributes C_1 (economic), C_2(functional), and C_3(operational), the weight vector of the attribute C_j $(j = 1, 2, 3)$ is $w = (0.5, 0.2, 0.3)^T$. Meanwhile, the attributes are all benefit attribute. Assume that the alternatives A_i $(i = 1, 2, 3, 4, 5)$ with respect to the attribute C_j $(j = 1, 2, 3)$ are represented by the single-valued neutrosophic number, as shown in Table 6.

Table 6: The alternatives with respect to the attribute denoted by SVNSs

	C_1	C_2	C_3
A_1	$< 0.0, 0.0, 0.0 >$	$< 1.0, 0.0, 0.0 >$	$< 0.0, 0.0, 1.0 >$
A_2	$< 0.0, 0.1, 0.1 >$	$< 0.9, 0.1, 0.2 >$	$< 0.4, 0.2, 0.2 >$
A_3	$< 0.1, 0.1, 0.1 >$	$< 0.9, 0.2, 0.3 >$	$< 0.5, 0.3, 0.2 >$
A_4	$< 0.3, 0.0, 0.1 >$	$< 1.0, 0.1, 0.2 >$	$< 0.4, 0.3, 0.2 >$
A_5	$< 0.0, 0.1, 0.0 >$	$< 0.9, 0.1, 0.2 >$	$< 0.4, 0.3, 0.2 >$

Table 7: A comparison studies with some exiting methods of Example 3

Method	The final ranking	The optimal alternative
Method from [29]	$A4 > A_1 > A_2 > A_5 > A_3$	A_4
Method from [30]	$A_4 > A_1 > A_5 > A_2 > A_3$	A_4
Method from [32]	N/A	failed
Method from [33]	N/A	failed
Method from [31]	$A_4 > A_2 > A_5 > A_1 > A_3$	A_4
Our method	$A_4 > A_1 > A_5 > A_2 > A_3$	A_4

By applying Equation (2), the similarity between an alternative A_i $(i = 1, 2, 3, 4, 5)$ and the ideal alternative $A^* = (< 1, 0, 0 >, < 1, 0, 0 >, < 1, 0, 0 >)$ is

as follows:

$$\vartheta^*(A_1, A^*)=\frac{1}{3}\Big[0.5\times(0+1+1)+0.2\times(1+1+1)+0.3\times(0+1+0)\Big]=0.6333,$$

$$\vartheta^*(A_2, A^*)=\frac{1}{3}\Big[0.5\times(0+\frac{243}{290}+\frac{243}{290})+0.2\times(\frac{171}{210}+\frac{243}{290}+\frac{208}{290})$$
$$+0.3\times(\frac{56}{210}+\frac{208}{290}+\frac{208}{290})\Big]=0.6074,$$

$$\vartheta^*(A_3, A^*)=0.5976, \qquad \vartheta*(A_4, A^*)=0.6769, \qquad \vartheta*(A_5, A^*)=0.6264$$

The results shows that the ranking order of the five alternatives is $A_4 > A_1 > A_5 > A_2 > A_3$. Therefore, the best supplier is A_4.

Comparative analysis of existing methods.

In order to further verify the practicability of the proposed new similarity measure, we use some exiting methods to calculate Example 3. The results can be seen in Table 7. By comparison, the practicability and superiority over our method is summarized as follows.

(1) The optimal alternative obtained by our method is the same as that in Refs. [29, 30, 31].

(2) Due to the division by zero problem, the ranking method of [32, 33] cannot apply to Example 3. But our method can be used to calculate it.

(3) The final ranking of the five alternatives obtained by our method is the same as using the Multi-Attributive Border Approximation area Comparison (MABAC) proposed by Peng [30]. However, the computational complexity of our method is much smaller than that in [30].

(4) Stability analysis. If the range of the weight vectors $w=(w_1, w_2, w_3)^T$ of the attribute is changed into $w_1 \in [0.4, 0.6], w_2 \in [0.1, 0.3]$, the optimal alternative calculated by our method is till A_4. This shows that our method is robust and stable.

5. Conclusion

Similarity measure (or distant measure) is an important tool for fuzzy decision-making and clustering. Under the single-valued neutrosophic set environment, the proposed similarity measure not only considers the relationship between the corresponding factors, but also considers the influence of the overall factors of the similarity. Then we use the new similarity measure in pathological diagnosis, clustering and decision-making to demonstrate its effectiveness. Comparing with the existing methods, it can be found that our method has the characteristics of high accuracy and wide application range.

Acknowledgment

This work is supported by National Natural Science Foundation Project (11701089, 11871259), the Key Program of the Natural Science Foundation of Fujian Province (2020J02043), Fujian Natural Science Foundation (2018J01422,2020J01801).

References

[1] L.A. Zadeh, *Fuzzy sets*, Information and Control, 8 (1965), 338-356.

[2] K. Atanassov, *Intuitionistic fuzzy sets*, Fuzzy sets and Systems, 20 (1986), 87-96.

[3] F. Smarandache, *A unifying field in logics: neutrosophy logic*, American Research Press: Rehoboth, Delaware, 1999, 1-141.

[4] H. Wang, F. Smarandache, Y.Q.Zhang, R. Sunderraman, *Single valued neutrosophic sets*, Review of the Air Force Academy, 4 (2010), 410-413.

[5] H. Garg, D. Rani, *A robust correlation coefficient measure of complex intuitionistic fuzzy sets and their applications in decision-making*, Applied Intelligence, 49 (2019), 496-512.

[6] J. Ye, *Intuitionistic fuzzy hybrid arithmetic and geometric aggregation operators for the decision-making of mechanical design schemes*, Applied Intelligence, 47 (2017), 743-751.

[7] I. Deli, Y. Şubaş, *A ranking method of single valued neutrosophic numbers and its applications to multi-attribute decision making problems*, International Journal of Machine Learning and Cybernetics, 8 (2017), 1309-1322.

[8] Z.S. Xu, *Intuitionistic fuzzy Bonferroni means*, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41 (2011), 568-578.

[9] V. Uluçay, A. Kılıç, I. Yıdız, M. Şahin, *A new approach for multi-attribute decision-making problems in bipolar neutrosophic sets*, Neutrosophic Set and Systems, 23 (2018), 142-159.

[10] Z.S. Xu, J. Chen, J.J.Wu. *Clustering, the algorithm is clearly adopted from Clustering algorithm for intuitionistic fuzzy sets*, Information Sciences, 178 (2008), 3775-3790.

[11] H.M. Zhang, Z.S. Xu, Q. Chen, *On clustering approach to intuitionistic fuzzy sets*, Control and Decision, 22 (2007),882-888.

[12] H. Verma, R.K. Agrawal, A. Sharan, *An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation*, Applied Soft Computing, 46 (2016), 543-557.

[13] Q.Y. Li, Y.C. Ma, F. Smarandache, S.W. Zhu, *Single-valued neutrosophic clustering algorithm based on Tsallis entropy maximization*, Axioms, 7 (2018), 57.

[14] M.N. Qureshi, M.V. Ahamad, *An improved method for image segmentation using k-means clustering with neutrosophic logic*, Procedia Computer Science, 132 (2018), 534-540.

[15] R.T. Ngan, M. Ali, H.S. Le, *δ-equality of intuitionistic fuzzy sets: a new proximity measure and applications in medical diagnosis*, Applied Intelligence, 48 (2018), 499-525.

[16] M. Luo, R. Zhao, *A distance measure between intuitionistic fuzzy sets and its application in medical diagnosis*, Artificial Intelligence in Medicine, 89 (2018), 34-39.

[17] Z.S. Xu. *On correlation measures of intuitionistic fuzzy sets*, Lecture Notes in Computer Science, 4224 (2006), 16-24.

[18] G. Wei, *Some cosine similarity measures for picture fuzzy sets and their applications to strategic decision making, Informatica*, 28 (2017), 547-564.

[19] V. Radhakrishna, S.A. Aljawarneh, P.V. Kumar, V. Janakie, *A novel fuzzy similarity measure and prevalence estimation approach for similarity profiled temporal association pattern mining*, Future Generation Computer Systems, 83 (2018), 582-595.

[20] Y.Tang, L.L. Wen, G.W. Wei, *Approaches to multiple attribute group decision making based on the generalized Dice similarity measures with intuitionistic fuzzy information*, International Journal of Knowledge-based and Intelligent Engineering Systems, 21 (2017), 85-95.

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

[22] C.M.Hwang, M.S. Yang, W.L. Hung, *New similarity measures of intuitionistic fuzzy sets based on the Jaccard index with its application to clustering*, International Journal of Intelligent Systems, 33 (2018), 1672-1688.

[23] F. Wang, J. Mao, *Aggregation similarity measure based on intuitionistic fuzzy closeness degree and its application to clustering analysis*, Journal of Intelligent & Fuzzy Systems, 35 (2018), 609-625.

[24] J. Ye, *Single-valued neutrosophic similarity measures based on cotangent function and their application in the fault diagnosis of steam turbine*, Soft Computing, 21 (2017), 817-825.

[25] J. Ye, *Single-valued neutrosophic clustering algorithms based on similarity measures*, Journal of Classification, 34 (2017), 148-162.

[26] W. Jiang, Y. Shou, *A novel single-valued neutrosophic set similarity measure and its application in multicriteria decision-making*, Symmetry, 9 (2017), 127.

[27] H.L. Huang, *New distance measure of single-valued neutrosophic sets and its application*, International Journal of Intelligent Systems, 31 (2016), 1021-1032.

[28] J. Ye, *Clustering methods using distance-based similarity measures of single-valued neutrosophic sets*, Journal of Intelligent Systems, 23 (2014), 379-389.

[29] P. Biswas, S. Pramanik, B.C. Giri, *TOPSIS method for multi-attribute group decision-making under single-valued neutrosophic environment*, Neural Computing and Applications, 27 (2016), 727-737.

[30] X. Peng, *Approaches to single-valued neutrosophic MADM based on MABAC, TOPSIS and new similarity measure with score function*, Neural Computing and Applications, 29 (2018), 939-954.

[31] J. Ye, *Improved correlation coefficients of single valued neutrosophic sets and interval neutrosophic sets for multiple attribute decision making*, Journal of Intelligent and Fuzzy Systems, 27 (2014), 2453-2462.

[32] S. Pramanik, P. Biswas, B.C. Giri, *Hybrid vector similarity measures and their applications to multi-attribute decision making under neutrosophic environment*, Neural Computing and Applications, 28 (2017), 1163-1176.

[33] J. Ye, *Multicriteria decision-making method using the correlation coefficient under single-value neutrosophic environment*, International Journal of General Systems, 42 (2013), 386-394.

[34] P. Majumdar, S.K. Samanta, *On similarity and entropy of neutrosophic sets*, Journal of Intelligent & Fuzzy Systems, 26 (2014), 1245-1252.

[35] V. Uluçay, I. Deli, M. Şahin, *Similarity measure of bipolar neutrosophic sets and their application to multiple criteria decision making*, Neural Computing and Applications, 29 (2016), 739-748.

[36] V. Uluçay, A. Kılıç, M. Şahin, H. Deniz, *A new hybrid distance-based similarity measure for refined neutrosophic sets and its application in medical diagnosis*, Matematika, 35 (2019), 83-96.

[37] M. Şahin, N. Olgun, V. Uluçay, A. Kargın, F. Smarandache, *A new similarity measure on falsity value between single valued neutrosophic sets based on the centroid points of transformed single valued neutrosophic numbers with applications to pattern recognition*, Neutrosophic Sets and Systems, 15 (2017), 31-48.

[38] W. Kempf, C. Mitteldof, *Pathologic diagnosis of cutaneous lymphomas*, Dermatologic Clinics, 33 (2015), 655-681.

[39] J. Zhang, Y.L. Song, et al. *Rapid and accurate intraoperative pathological diagnosis by artificial intelligence with deep learning technology*, Medical Hypotheses, 107 (2017), 98-99.

[40] M.J.P. Castanho, F. Hernandes, A.M. De Réb, S. Rautenbergb, A. Billisc, *Fuzzy expert system for predicting pathological stage of prostate cancer*, Expert Systems with Applications, 40 (2013), 466-470.

[41] S. Kolhe, R. Kamal, H.S.Sainid, G.K. Guptab, *A web-based intelligent disease-diagnosis system unsing a new fuzzy-logic based approach for drawing the inferences in crops*, Computers and Electronics in Agriculture, 76 (2011), 16-27.

[42] S.R. Chowdhury, H. Saha, *Development of a FPGA based fuzzy neural network system for early diagnosis of critical health condition of a patient*, Computers in Biology and Medicine, 40 (2010), 190-200.

Accepted: March 11, 2020